Programme

TUT: Tutorial, KL: Keynote Lecture, IS: Invited Session, TT: Thematic Track,
CS: Contributed Session, PS: Poster Session

Tuesday, 19 July

8:30–9:30 **Registration**

9:30-11:00 **Room: 118**

TUT **Tutorial 1** [T1]
282 Analysis of Data Streams
João Gama

11:00–11:30 **Coffee Break**

11:30-13:00 **Room: 118**

TUT **Tutorial 1** [T1]
282 Analysis of Data Streams
João Gama

13:00–14:30 **Lunch Break**

14:30-16:00 **Room: 118**

TUT **Tutorial 2** [T2]
283 Categorical Data Analysis and Visualization
Rosaria Lombardo and Eric Beh

16:00–16:30 **Coffee Break**

16:30-18:00 **Room: 118**

TUT **Tutorial 2** [T2]
283 Categorical Data Analysis and Visualization
Rosaria Lombardo and Eric Beh

18:30-20:30 **Ice Breaker** *House of Vinho Verde*

Wednesday, 20 July

8:00–9:00 **Registration**

9:00-9:30 **Room: Grand Hall**

OS **Opening Session**

9:30-10:30 **Room: Grand Hall**

KL **Keynote Lecture: Dianne Cook** [K1]

Chair: Angela Montanari

256 A Showcase of New Methods for High Dimensional Data Viewing with Linear Projections and Sections
Dianne Cook

10:30–11:00 **Coffee Break**

11:00-12:00 **Room: 260**

TT **Thematic Track: Fifty Years of Biplots 1** [BIP1]

Chair: Niël le Roux

240 Biplots for Categorical and Ordinal Data Based on Logistic Responses
Jose Luis Vicente-Villardon

75 The Biplot Inner Product for Interpretation and Derivation of Eigenvector Methods
Cajo ter Braak

158 Fifty Years of Biplots: Some Remaining Enigmas and Challenges
Jan Graffelman

11:00-12:00 **Room: 113**

TT **Thematic Track: Functional Data Analysis 1** [FDA1]

Chair: Michelle Carey

46 Outlier Detection for BIG Functional Data
Rosa E. Lillo, Oluwasegun T. Ojo, and Antonio Fernández-Anta

85 Outlier and Novelty Detection for Functional Data: a Semiparametric Bayesian Approach
Francesco Denti, Andrea Cappozzo, and Francesca Greselin

235 A Geometric Perspective on Functional Outlier Detection
Moritz Herrmann and Fabian Scheipl

Wednesday, 20 July

11:00-12:00 **Room: 118**

TT **Thematic Track: Model-based Clustering 1** [MBC1]

Chair: Salvatore Ingrassia

6 A New Decomposition of Orthogonal Matrices with Application to Common Principal Components
Luca Bagnato and Antonio Punzo

18 An MML Embedded Approach for Estimating the Number of Clusters
Cláudia Silvestre, Margarida Cardoso, and Mário Figueiredo

76 Comparison of Segmentation Approaches for Partial Least Squares Path Modeling with Stability Assessment
Sophie Dominique, Mohamed Hanafi, Fabien Llobell, Jean-Marc Ferrandi, and Véronique Cariou

11:00-12:00 **Room: 159**

TT **Thematic Track: Robust Methods 1** [RM1]

Chair: Luis Ángel García-Escudero

232 Robust Classification for Toroidal Data
Giovanni Saraceno, Luca Greco, and Claudio Agostinelli

138 Consistency of Trimmed Estimators of Scatter Under the t-distribution
Andrea Cerioli, Lucio Barabesi, Luis Ángel García-Escudero, and Agustín Mayo-Íscar

148 Robust Classification in High Dimensions Using Regularized Covariance Estimates
Valentin Todorov and Peter Filzmoser

11:00-12:00 **Room: 115**

TT **Thematic Track: Symbolic Data Analysis 1** [SDA1]

Chair: Rosanna Verde

157 Symbolic Concordance and Discordance Illustrated on Data from an International Teaching and Learning Survey
Simona Korenjak-Černe, Barbara Japelj Pavešić, and Edwin Diday

275 The Use of Regression to Partition a Dataset of Interval Observations
Lynne Billard and Fei Liu

187 A Clusterwise Regression Method for Distributional Data
Rosanna Verde, Antonio Balzanella, and Antonio Irpino

11:00-12:00 **Room: 211**

TT **Thematic Track: Statistical Learning and Data Mining 1** [SLDM1]

Chair: Jangsun Baek

99 Heterogeneous Random Forests
Ye-eun Kim and Hyunjoong Kim

105 Analysis of the Damage Rate Using Typhoon Information
Su Hoon Choi and Min Soo Kim

144 Resampling, Relabeling, and Raking for Extremely Imbalanced Classification
Hae-Hwan Lee, Seunghwan Park, and Jongho Im

Wednesday, 20 July

11:00-12:00 **Room: 213**

TT **Thematic Track: Social Network Analysis 1** [SNA1]

Chair: Vladimir Batagelj

38 Clustering Student Mobility Data in 3-way Networks
Vincenzo Giuseppe Genova, Giuseppe Giordano, Giancarlo Ragozini, and Maria Prosperina Vitale

195 Multi-perspective Risky User Classification in Social Networks
Antonio Pellicani, Gianvito Pio, and Michelangelo Ceci

13 Clustering and Blockmodeling Temporal Networks - Two Indirect Approaches
Vladimir Batagelj

11:00-12:00 **Room: 135**

CS **Contributed Session: Clustering 1** [CL1]

Chair: Nema Dean

50 Clustering Validation in the Context of Hierarchical Cluster Analysis: an Empirical Study
Osvaldo Dias Lopes da Silva, Áurea Sandra Toledo de Sousa, and Helena Bacelar-Nicolau

160 Divide and Conquer: a Clustering Method for Hierarchical and Nested Data Structures
Andrej Svetlosak, Miguel de Carvalho, Gabriel Martos Venturini, and Raffaella Calabrese

237 Significance Mode Analysis (SigMA) for Hierarchical Structures
Sebastian Ratzenböck, Torsten Möller, Josefa E. Großschedl, João Alves, Immanuel M. Bomze, and Stefan Meingast

11:00-12:00 **Room: 218**

CS **Contributed Session: Machine Learning 1** [ML1]

Chair: Miguel de Carvalho

129 Kurtosis-based Projection Pursuit for Matrix-valued Data
Una Radojicic, Klaus Nordhausen, and Joni Virta

154 Comparison of Pixel Based Segmentation Methods in Papillary Thyroid US Images
Neslihan Gökmen İnan, İsmail Meşe, Düzgün Yıldırım, and Ozan Kocadağlı

173 Bootstrapping Binary GEV Regressions for Massive Unbalanced Datasets
Michele La Rocca, Marcella Niglio, and Marialuisa Restaino

Wednesday, 20 July

11:00-11:40 **Room: 215**

CS **Contributed Session: Supervised Machine Learning** [S-ML]

Chair: Taesung Park

40 Supervised Classification via Neural Networks for Replicated Point Patterns
Kateřina Pawlasová, Iva Karafiátová, and Jiří Dvořák

98 Covariate Selection Method in Propensity Score Model for the Quantile Treatment Effect Estimation
Takehiro Shoji, Jun Tsuchida, and Hiroshi Yadohisa

12:00-13:00 **Room: 213**

TT **Thematic Track: Data Science in Social Sciences 1** [DS-SS1]

Chair: Theodore Chadjipadelis

91 Are Attitudes Toward Immigration Changing in Europe? An Analysis Based on Latent Class IRT Models
Ewa Genge and Francesco Bartolucci

153 Visualization of IATA Regions in Air Transport Before and After the COVID-19 Pandemic
Tüzün Tolga İnan, Neslihan Gökmen İnan, Aylin Yaman Kocadağlı, and Ozan Kocadağlı

33 Political and Religious Attitudes in Greece: Behavioral Discourses
Georgia Panagiotidou and Theodore Chadjipadelis

12:00-13:00 **Room: 113**

TT **Thematic Track: Functional Data Analysis 2** [FDA2]

Chair: Andrea Cappozzo

84 Functional Data Representation with Merge Trees
Matteo Pegoraro and Piercesare Secchi

135 Elastic Regression for Irregularly Sampled Curves in $\mathbb{R}^d$
Lisa Steyer, Almond Stöcker, and Sonja Greven

226 Misalignment of Spectral Data: Constrained Optimization in a Functional Data Analysis Framework
Francesca Di Salvo, Delia Francesca, Chillura Martino, and Gabriella Chirco

Wednesday, 20 July

12:00-13:00 **Room: 118**

TT **Thematic Track: Model-based Clustering 2** [MBC2]

Chair: Francesca Greselin

24 Model Based Clustering of Functional Data with Mild Outliers
Cristina Anton and Iain Smith

55 Old and New Constraints in Model Based Clustering
Luis Ángel García Escudero, Agustín Mayo-Íscar, Gianluca Morelli, and Marco Riani

79 Model Based Clustering and Outlier Detection with Missing Data
Cristina Tortora, Hung Tong, and Louis Tran

12:00-13:00 **Room: 159**

TT **Thematic Track: Robust Methods 2** [RM2]

Chair: Luis Ángel García-Escudero

90 How to Mitigate the Effect of Outliers on Balancing Technique
Rasool Taban, Maria do Rosário Oliveira, and Cláudia Nunes Philippart

114 Outlier Detection in Functional Data
Martial Amovin-Assagba, Irène Gannaz, and Julien Jacques

132 Robustified Elastic Net Estimator for Multinomial Regression
Fatma Sevinc Kurnaz and Peter Filzmoser

12:00-13:00 **Room: 115**

TT **Thematic Track: Symbolic Data Analysis 2** [SDA2]

Chair: Rosanna Verde

56 Optimized Symbolic Correspondence Analysis for Multi-valued Variables
Jorge Andres Arce Garro and Oldemar Rodriguez Rojas

59 Symbolic t-SNE and UMAP Methods for Interval Type Variables
Oldemar Rodriguez Rojas

86 Two-stage Principal Component Analyses on Interval-valued Data Using Patterned Covariance Structure
Anuradha Roy

12:00-13:00 **Room: 211**

TT **Thematic Track: Statistical Learning and Data Mining 2** [SLDM2]

Chair: Hyunjoong Kim

27 Detection of the Biliary Atresia Using Deep Convolutional Neural Networks Based on Statistical Learning Weights via Optimal Similarity and Resampling Methods
Kuniyoshi Hayashi, Eri Hoshino, Mitsuyoshi Suzuki, Erika Nakanishi, Kotomi Sakai, and Masayuki Obatake

131 Variational Autoencoder with Gamma Mixture for Clustering Right-skewed Data
Jinwon Heo and Jangsun Baek

133 An Efficient Way to Identify Inliers via Inlier-memorization Effect of Deep Generative Models
Dongha Kim, Jaesung Hwang, and Yongdai Kim

Wednesday, 20 July

12:00-13:00 **Room: 135**

CS **Contributed Session: Clustering 2** [CL2]

Chair: Mohamed Nadif

10 Three-way Spectral Clustering
Cinzia Di Nuzzo and Salvatore Ingrassia

25 Fuzzy Clustering by Hyperbolic Smoothing
David Masis, Esteban Segura, Javier Trejos, and Adilson E. Xavier

184 Combining KDE and DBSCAN Clustering to Understand Road Traffic Accidents: The Case of Setúbal, Portugal
Pedro Nogueira, Marcelo Silva, Paulo Infante, Paulo Rebelo Manuel, Leonor Rego, Anabela Afonso, and Gonçalo Jacinto

12:00-13:00 **Room: 218**

CS **Contributed Session: Machine Learning 2** [ML2]

Chair: Adalbert Wilhelm

11 Similarity Forest for Time Series Classification
Tomasz Górecki, Maciej Łuczak, and Paweł Piasecki

172 Uncovering Regions of Maximum Dissimilarity on Random Process Data
Miguel de Carvalho and Gabriel Martos

189 Franz Liszt's Transcendental Etudes: an Evolutionary Analysis by Machine Learning
Matteo Farnè

12:00-13:00 **Room: 260**

CS **Contributed Session: Supervised Classification 1** [SC1]

Chair: Matthijs Warrens

126 Quantile-distribution Functions and Their Use for Classification
Edoardo Redivo, Cinzia Viroli, and Alessio Farcomeni

205 Analysis of Gini Splitting Criterion and Comparison with Maximum Likelihood Rule
Amirah Alharthi and Charles Taylor

230 Envelope-based Support Vector Machine
Alya Alzahrani and Andreas Artemiou

12:00-13:00 **Room: 215**

CS **Contributed Session: Statistical and Econometric Methods** [SEM]

Chair: Herbert Lee

61 A Moment-free Measure of Multivariate Skewness
Andrzej Sokołowski and Małgorzata Markowska

97 The Weighted RV Coefficient: Exact Moments by Invariant Orthogonal Integration
François Bavaud

141 Testing Equality of Multivariate Coefficients of Variation
Marc Ditzhaus and Łukasz Smaga

13:00–14:30 **Lunch Break**

Wednesday, 20 July

14:30-16:00 **Room: 118**

IS **Invited Session: Classification of Time Series** [INV1]

Chair: José A. Vilar Fernández

81 A Criterion for Selecting the Number of Time Series Clusters
Daniel Peña and Ruey Tsay

149 Clusters Based on Prediction Accuracy of Global Forecasting Models
Pablo Montero Manso, Angel Lopez-Oriona, and José Vilar

167 Clustering and Classifying Time Series in the Sktime Toolkit: a Practical Review of Latest Advances in the Field
Anthony Bagnall

14:30-16:00 **Room: 218**

IS **Invited Session: Benchmarking Challenge** [BC]

Chair: Christian Hennig

264 Vine Copula Mixture Models and Clustering for Non-Gaussian Data
Özge Sahin and Claudia Czado

273 Pitfalls of Automatic Optimization Procedures and Benchmarking in Cluster Analysis
Quirin Stier and Michael Thrun

265 Comparison of Similarity Measures for Categorical Data in Hierarchical Clustering
Zdeněk Šulc and Hana Rezankova

268 Comparing Model Selection Techniques to Determine the Number of Overlapping Clusters for the Additive Profile Clustering Model
Tom Frans Wilderjans, Julian Rossbroich, and Jeffrey Durieux

16:00–16:30 **Coffee Break**

16:30-17:30 **Room: 215**

TT **Thematic Track: Compositional Data Analysis 1** [CoDA1]

Chair: Glòria Mateu-Figueras

14 A New Regression Model for the Analysis of Microbiome Data
Roberto Ascari and Sonia Migliorati

44 The Death Process in Italy Before and During the Covid-19 Pandemic: a Functional Compositional Approach
Riccardo Scimone, Alessandra Menafoglio, Laura Sangalli, and Piercesare Secchi

201 Sampling Design for Uncovering Natural Laws in Compositional Data
Lan Liang, Glòria Mateu-Figueras, and Jan Graffelman

Wednesday, 20 July

16:30-17:50 **Room: 113**

TT **Thematic Track: Functional Data Analysis** [FDA3]

Chair: Alessia Pini

21 Penalized Model-based Functional Clustering: a Regularization Approach via Shrinkage Methods
Nicola Pronello, Rosaria Ignaccolo, Luigi Ippoliti, and Sara Fontanella

47 Clustering in FDA Mixing the Epigraph and the Hypograph Indexes with Machine Learning Algorithms
Belén Pulido Bravo, Alba María Franco Pereira, and Rosa Elvira Lillo Rodríguez

152 Localization Processes for Functional Data Classification
Antonio Elías, Raúl Jiménez, and Joseph E. Yukich

177 A New Functional Data Clustering Technique Based on Spectral Clustering and Downsampling
Maryam Al Alawi, Surajit Ray, and Mayetri Gupta

16:30-17:50 **Room: 118**

TT **Thematic Track: Model-based Clustering 3** [MBC3]

Chair: Andrea Cerioli

28 Mixtures of Seemingly Unrelated Contaminated Normal Regression Models
Gabriele Perrone and Gabriele Soffritti

257 Monitoring Hyperparameter Choice for Robust Cluster Weighted Model
Francesca Greselin, Andrea Cappozzo, Luis Ángel García Escudero, and Agustín Mayo-Íscar

198 Reinforced EM Algorithm Through Clever Initialization for Clustering with Gaussian Mixture Models
Joshua Tobin, Chin Pang Ho, and Mimi Zhang

32 Latent Block Regression Model
Rafika Boutalbi, Lazhar Labiod, and Mohamed Nadif

16:30-17:30 **Room: 135**

CS **Contributed Session: Clustering 3** [CL3]

Chair: Christian Hennig

37 Towards a Bi-stochastic Matrix Approximation of k-means and Some Variants
Lazhar Labiod and Mohamed Nadif

34 Clustering Brain Connectomes Through a Density-peak Approach
Riccardo Giubilei

41 New Metrics for Classifying Phylogenetic Trees Using k-means and the Symmetric Difference Metric
Nadia Tahiri and Aleksandr Koshkarov

Wednesday, 20 July

16:30-17:50 **Room: 218**

CS **Contributed Session: Dimension Reduction** [DR]

Chair: Ndèye Niang

71 Alternating Optimization Framework for Sparse Simultaneous Component Analysis Based on Data Integration
Rosember Guerra-Urzola, Juan C. Vera, Katrijn Van Deun, and Klaas Sijtsma

214 Joint Sparse PCA
Katrijn Van Deun

102 Joint Sparse Principal Component Analysis: a Simulation Study
Tra Le and Katrijn Van Deun

219 Copula-based Non-metric Unfolding on Augmented Data Matrix
Marta Nai Ruscone and Antonio D'Ambrosio

16:30-17:50 **Room: 211**

CS **Contributed Session: Data Science in Health Sciences** [DS-HS]

Chair: Utkarsh Dang

31 Emotion Classification Based on Single Electrode Brain Data: Applications for Assistive Technology
Duarte Rodrigues, Luis Paulo Reis, and Brígida Mónica Faria

125 On the Role of Data, Statistics and Decisions in a Pandemic
Ursula Garczarek, Beate Jahn, Sarah Friedrich, Joachim Behnke, Joachim Engel, Ralf Muennich, Markus Pauly, Adalbert Wilhelm, Olaf Wolkenhauer, Markus Zwick, Uwe Siebert, and Tim Friede

254 A Deep Learning Analytics to Detect Dental Caries
Taerim Lee

255 Identification of Shared Genetic Loci Between Psychiatric Disorders and Telomere Length and Evaluation of Their Role as Potential Drug Targets
Claudia Pisanu, Anna Meloni, and Alessio Squassina

16:30-17:50 **Room: 260**

CS **Contributed Session: Supervised Classification - Trees** [SCT]

Chair: Charles Taylor

82 Estimating Optimal Decision Trees for Treatment Assignment with $k > 2$ Treatment Alternatives: a Classification Problem with a Unit- and Class-dependent Misclassification Cost
Iven Van Mechelen and Aniek Sies

210 ExactTree: an R-package for Globally Optimal Decision Trees
Elise Dusseldorp, Juan Claramuny Gonzales, Jacqueline Meulman, Samil Uysal, and Bart Jan van Os

239 Optimal Random Projection Trees Ensemble
Nosheen Faiz, Adi Lausen, Metodi Metodiev, Zardad Khan, and Berthold Lausen

242 Born-again and Bayesian Methods for Improving the Performance of Decision Trees
Marjolein Fokkema

Wednesday, 20 July

16:30-17:50 **Room: 213**

CS **Contributed Session: Text Mining** [TM]

Chair: Jörg Blasius

30 Evolution of Media Coverage on Climate Change and Environmental Awareness: an Analysis of Tweets from UK and US Newspapers
Gianpaolo Zammarchi, Maurizio Romano, and Claudio Conversano

48 Improving Classification of Documents by Semi-supervised Clustering in a Semantic Space
Jasminka Dobša and Henk A.L. Kiers

168 Is It Hate or Criticism? An Exploratory Approach to Negative Comments on YouTube
Manuela Schmidt

171 A Time-Varying Text Based Ideal Point Model to Infer Partisanship in the U.S. Senate
Sourav Adhikari, Bettina Gruen, and Paul Hofmarcher

16:30-17:50 **Room: 115**

CS **Contributed Session: Time Series** [TS]

Chair: Daniel Peña

7 Oracle-LSTM: a Neural Network Approach to Mixed Frequency Timeseries Prediction
Alessandro Bitetto and Paola Cerchiello

109 Time Series of Counts Under Censoring
Isabel Silva, Maria Eduarda Silva, Isabel Pereira, and Brendan McCabe

164 Multivariate Time Series Feature Extraction via Multilayer Networks
Vanessa Silva, Maria Eduarda Silva, Pedro Ribeiro, and Fernando Silva

258 On the Use of the Choquet Fuzzy Integral to Aggregate Predictions of Time Series: an Application to Economic (and Other Types of) Data
Diogo Alves, José Matos, and Sandra Silva

19:00-22:00 **Welcome Reception** *Praia da Luz*

Thursday, 21 July

8:00–9:00 **Registration**

9:00-10:00 **Room: 260**

TT **Thematic Track: Fifty Years of Biplots 2** [BIP2]

Chair: Jose Luis Vicente-Villardon

96 Some Biplot Alternatives
Patrick Groenen

222 Biplots in Dimension Reduction and Clustering
Alfonso Iodice D'Enza, Angelos Markos, and Michel van de Velden

145 Biplots: A Sophisticated Multivariate Approach or a User-Friendly Technique?
Manuel Rui Alves

9:00-9:40 **Room: 215**

TT **Thematic Track: Compositional Data Analysis 2** [CoDA2]

Chair: Jan Graffelman

170 PLS-based Principal Balances for Regression and Classification with High-dimensional Compositional Data
Viktorie Nesrstová, Ines Wilms, Karel Hron, Josep Antonio Martín-Fernández, Peter Filzmoser, and Javier Palarea-Albaladejo

174 Clustering Count Data Using Compositional Methods
Marc Comas-Cufí, Josep A. Martín-Fernández, Glòria Mateu-Figueras, and Javier Palarea-Albaladejo

9:00-10:00 **Room: 159**

TT **Thematic Track: Data Science in Economics and Finance 1** [DS-EF1]

Chair: Krzysztof Jajuga

121 Urban Development Paths in Poland: Multidimensional Perspective
Jacek Batóg and Barbara Batóg

139 Analyzing the Evolution of EU Countries and Indicators of Europe 2020 Agenda
Adelaide Figueiredo and Fernanda Otília Figueiredo

185 Google Trends as a Macroeconomic Predictor: Behind the Scenes
Eduardo André Costa and Maria Eduarda Silva

9:00-10:00 **Room: 213**

TT **Thematic Track: Data Science in Social Sciences 2** [DS-SS2]

Chair: Alice Barth

49 COVID-19 Pandemic: a Methodological Model for the Analysis of Government's Preventing Measures and Health Data Records
Theodore Chadjipadelis and Sofia Magopoulou

113 Detecting Fabricated Interviews Using the Hamming Distance
Jörg Blasius

180 Digital Development and Internet Use in the European Union Countries
Fernanda Otília Figueiredo and Adelaide Maria Figueiredo

Thursday, 21 July

9:00-10:00 **Room: 113**

TT **Thematic Track: Functional Data Analysis 4** [FDA4]

Chair: Rosaria Ignaccolo

78 Probabilistic Clustering with Local Alignment of Italian COVID-19 Death Curves
Marzia Cremona, Tobia Boschi, and Francesca Chiaromonte

155 Model Free Predictive Inference for Functional Kriging Techniques Based on Conformal Prediction
Andrea Diana, Elvira Romano, and Jorge Mateu

212 Density Modelling via Functional Data Analysis
Stefano Antonio Gattone and Tonio Di Battista

9:00-10:00 **Room: 118**

TT **Thematic Track: Model-based Clustering 4** [MBC4]

Chair: Hien Nguyen

19 On Parsimonious Modelling via Matrix-variate t Mixtures
Salvatore Daniele Tomarchio

123 Four Skewed Tensor Variate Distributions
Michael Gallaugher, Peter Tait, and Paul McNicholas

204 A Family of Skewed Power Exponential Mixture Models for Clustering and Classification
Utkarsh Dang, Michael Gallaugher, Ryan Browne, and Paul McNicholas

9:00-10:00 **Room: 211**

TT **Thematic Track: Optimization in Classification and Clustering 1** [OCC1]

Chair: Laura Palagi

77 Generating Collective Counterfactual Explanations in Score-Based Classification via Mathematical Optimization
Jasone Ramírez-Ayerbe, Emilio Carrizosa, and Dolores Romero Morales

127 Spherical Separation in Machine Learning
Matteo Avolio, Annabella Astorino, and Antonio Fuduli

231 Model Extraction Based on Counterfactual Explanations
Cecilia Salvatore and Veronica Piccialli

Thursday, 21 July

9:00-10:00 **Room: 115**

TT **Thematic Track: Symbolic Data Analysis 3** [SDA3]

Chair: Javier Arroyo

107 Isolation Forests for Symbolic Data as a Tool for Outlier Mining
Marcin Pełka and Andrzej Dudek

119 Symbolic Clustering Methods Applied to Interval Estimates of Production Cost Quantiles
Dominique Desbois

209 Fisher Discriminant Analysis for Interval Data
Diogo Pinheiro, M. Rosário Oliveira, Igor Kravchenko, and Lina Oliveira

9:00-10:00 **Room: 135**

CS **Contributed Session: Clustering 4** [CL4]

Chair: Margarida Cardoso

29 Stability of Mixed-type Cluster Partitions for Determination of the Number of Clusters
Rabea Aschenbruck, Gero Szepannek, and Adalbert F. X. Wilhelm

192 Multinomial Multilevel Models with Discrete Random Effects: a Multivariate Clustering Tool
Chiara Masci, Francesca Ieva, and Anna Maria Paganoni

221 PD-Clustering for Mixed Data
Francesco Palumbo and Cristina Tortora

9:00-10:00 **Room: 218**

CS **Contributed Session: Machine Learning 3** [ML3]

Chair: Dongha Kim

2 Anomaly Detection-based Under-sampling for Imbalanced Classification Problems
You-Jin Park, Chun-Yang Peng, Rong Pan, and Douglas C. Montgomery

112 Continuous Adaptation to Distribution Drifts Through Continual Learning in Manufacturing
Henrique Siqueira and Onay Urfalioglu

270 Detecting Anomalies with TADGAN: a Case Study
Inês Oliveira e Silva, Carlos Soares, Arlete Rodrigues, and Pedro Bastardo

Thursday, 21 July

10:00-10:30 **Room: Main Hall, Ground Floor**

PS **Poster Session 1** [PS1]

45 A Trivariate Geometric Classification of Decision Boundaries for Mixtures of Regressions
Filippo Antonazzo and Salvatore Ingrassia

54 Clustering Rainfall by Simulated Annealing for Histogram Symbolic Data
Alejandro Chacón and Javier Trejos

57 Statistical Assessment of Youth Inclusion in the National Labour Markets
Beata Bal-Domańska

58 Barriers to Industry Digitization in Poland from the Perspective of High and Medium-high Technology Sector Enterprises
Elżbieta Sobczak, Marcin Pełka, and Karolina Pokorska

67 Kernel Smoothing-based Probability Contours for Tumour Segmentation
Wenhui Zhang and Surajit Ray

73 Genomic Prediction Using Machine Learning Methods: Performance Comparison on Synthetic and Empirical Data
Vanda Lourenço, Joseph Ogutu, Rui Rodrigues, and Hans-Peter Piepho

88 Pooled Mean and Confidence Interval Estimation Combining Different Sets of Summary Statistics
Flora Ferreira, José Soares, Fernanda Sousa, Filipe Magalhães, Isabel Ribeiro, and Pedro Pacheco

106 Prediction of Diabetes via Bayesian Network Classifier from Exposure to Environmental Polluting Chemicals Data
Rosy Oh, Hong Kyu Lee, Youngmi Pak, and Man-Suk Oh

137 Model Performance Metrics for Sample Selection Bias Correction by Pseudo Weighting
An-Chiao Liu, Ton de Waal, Katrijn Van Deun, and Sander Scholtus

140 Some Factors That Influence the Nature of Road Traffic Accidents
Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Vitor Nogueira, Paulo Quaresma, José Saias, Daniel Santos, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel

176 Comparing Variable Selection Methods for High-dimensional Compositional Data in a Discriminant Analysis Context
Pepus Daunis-i-Estadella, Glòria Mateu-Figueras, Viktorie Nesrstová, Karel Hron, and Josep Antoni Martín-Fernández

178 Cluster Analysis and Genetic Risk Score in Age-related Macular Degeneration - the Coimbra Eye Study
Rita Coimbra

183 Sensor System for Standardizing Articulation Patterns According to Korean Phonemes
Seong Tak Woo and Da Hee Oh

218 Trade and Bank Credit of Portuguese SMEs: a Panel Data Application
Carla Henriques, Pedro Pinto, and Carolina Cardoso

Thursday, 21 July

10:30–11:00 **Coffee Break**

11:00–11:15 **Room: Grand Hall**

BCA **Benchmarking Challenge Award**

Chair: Matthijs Warrens

11:15–12:00 **Room: Grand Hall**

AS **IFCS Awards**

Chair: Maurizio Vichi

Student/Postdoctoral Fellow Paper Competition and Travel Award
Student Poster Session Competition Award on Innovative Research Including Applications
Chikio Hayashi Award
Helga and Wolfgang Gaul Stiftung Award
IFCS Research Medal for Outstanding Research

12:00–13:00 **Room: Grand Hall**

KL **Keynote Lecture: Edwin Diday** [IFCS 2022 Research Medal Lecture]

Chair: Maurizio Vichi

281 An Introduction to S-concordance and S-discordance
Edwin Diday

13:00–14:30 **Lunch Break**

14:30–15:30 **Room: Grand Hall**

KL **Keynote Lecture: Charles Bouveyron** [K2]

Chair: José G. Dias

272 Statistical Learning with Dynamic Interaction Data for Public Health
Charles Bouveyron

16:30-20:30 **Excursion** *Boat Cruise on the Douro*

21:00 **Council Diner**

Friday, 22 July

8:30–9:00 **Registration**

9:00–10:30 **Room: 118**

IS **Invited Session: COVID Data Analysis** [INV2]

Chair: Taerim Lee

206 Effect of Type 2 Diabetes and Its Genetic Susceptibility on Severity and Mortality of COVID-19 in UK Biobank
Aeyeon Lee, Youngkwang Cho, Jun Li, Taesung Park, Wonil Chung, and Liming Liang

260 Prognosis of COVID-19 Patients by the Underlying Diseases and Drug Treatment in Korea
Ho Kim and Taerim Lee

262 Algorithms for Clustering COVID-19 Data: a Holistic Overview of Current Trends and New Visual Approaches
Eun-Kyung Lee

9:00–10:30 **Room: 218**

IS **Invited Session: Categorical Data Analysis and Visualization** [INV3]

Chair: Rosaria Lombardo

100 Embedded Word MCA Biplots for Sentiment Visualisation: Application to COVID-19 Related Tweets
Zoë-Mae Adams, Johané Nienkemper-Swanepoel, Niël le Roux, and Sugnet Lubbe

117 A General Framework for Implementing Distance Measures for Categorical Variables
Carlo Cavicchia, Michel van de Velden, Alfonso Iodice D'Enza, and Angelos Markos

151 Some Descriptive Statistics of Aggregated Symbolic Data
Junji Nakano, Nobuo Shimizu, and Yoshikazu Yamamoto

10:30–11:00 **Coffee Break**

11:00–12:00 **Room: Grand Hall**

KL **Keynote Lecture: Angela Montanari** [Presidential Address]

Chair: Paula Brito

284 Perturb and Conquer. How Classification Can Benefit from Data Perturbation
Angela Montanari

12:00–13:00 **Room: 260**

TT **Thematic Track: Fifty Years of Biplots 3** [BIP3]

Chair: Patrick Groenen

103 Categorical Data Visualization and the Cressie-Read Divergence Statistic
Eric Beh and Rosaria Lombardo

83 Biplots Based on Latent Variable Models in the Analysis of Ecological Communities
Jenni Niku and Sara Taskinen

251 Biplot Representation of Partial Least Squares Regression for Binary Responses
Laura Vicente-Gonzalez and Jose Luis Vicente-Villardon

Friday, 22 July

12:00–13:00 **Room: 113**

TT **Thematic Track: Functional Data Analysis 5** [FDA5]

Chair: Pedro Delicado

43 Generalized Spatio-temporal Regression with PDE Penalization
Eleonora Arnone, Elia Cunial, and Laura M. Sangalli

74 Impact Point Selection in Semiparametric Bifunctional Models
Silvia Novo, Germán Aneiros, and Philippe Vieu

134 Latent Function-on-Scalar Regression Models for Observed Sequences of Correlated Binary Data: a Restricted Likelihood Approach
Fatemeh Asgari and Valeria Vitelli

12:00–13:00 **Room: 211**

TT **Thematic Track: Optimization in Classification and Clustering 2** [OCC2]

Chair: Immanuel Bomze

15 pcTVI: Parallel MDP Solver Using a Decomposition Into Independent Chains
Jaël Champagne Gareau, Éric Beaudry, and Vladimir Makarenkov

143 Classification of Viral Pneumonia Images via Multiple Instance Learning
Antonio Fuduli, Matteo Avolio, Eugenio Vocaturo, and Ester Zumpano

188 Nonlinear Approaches for Multiple Instance Learning
Annabella Astorino, Matteo Avolio, and Antonio Fuduli

12:00–12:40 **Room: 118**

TT **Thematic Track: Model-based Clustering 5** [MBC5]

Chair: Sanjeena Dang

26 An Online Minorization-Maximization Algorithm
Hien Nguyen, Florence Forbes, Gersende Fort, and Olivier Cappe

93 Frugal Gaussian Clustering of Huge Imbalanced Datasets Through a Bin-marginal Approach
Filippo Antonazzo, Christophe Biernacki, and Christine Keribin

12:00–13:00 **Room: 159**

TT **Thematic Track: Robust Methods 3** [RM3]

Chair: Agustín Mayo Íscar

208 Exact Computation of the Angular Halfspace Depth
Stanislav Nagy and Rainer Dyckerhoff

217 Reconstruction of Atomic Measure Based on Its Simplicial Depth
Petra Laketa and Stanislav Nagy

5 Optimized Centroids as Robust Tools for Object Localization in Images
Jan Kalina and Patrik Janacek

Friday, 22 July

12:00–13:00 **Room: 115**

TT **Thematic Track: Symbolic Data Analysis 4** [SDA4]

Chair: Simona Korenjak Černe

108 Analysis of the Changes in the Polish Traditional Drugstores Market During COVID-19
Marcin Pełka, Antonio Irpino, and Michał Swachta

147 Logistic Regression Models for Aggregated Data
Thomas Whitaker, Boris Beranger, and Scott Sisson

186 Nonparametric Regressions for Distributional Data
Albert Meco, Javier Arroyo, and Antonio Irpino

12:00–13:00 **Room: 213**

TT **Thematic Track: Spatial-Temporal Data Analysis 1** [STDA1]

Chair: Koji Kurihara

122 Hotspot Cluster Detection Based on Spatial Hierarchical Structure and Its Software
Fumio Ishioka, Shoji Kajinishi, and Koji Kurihara

181 Group Lasso Penalty for Spatial Clustered Coefficient Regression
Toshiki Sakai, Jun Tsuchida, and Hiroshi Yadohisa

267 Visualization of the Number of New Positives for COVID-19 in Japan
Yoshiro Yamamoto, Sanetoshi Yamada, Mayumi Tanahashi, and Tadashi Imanishi

12:00–13:00 **Room: 135**

CS **Contributed Session: Clustering 5** [CL5]

Chair: Jean Diatta

95 Clustering with Missing Data: Which Imputation Model for Which Cluster Analysis Method?
Vincent Audigier, Ndèye Niang, and Matthieu Resche-Rigon

64 Two Simple but Efficient Algorithms to Recognize Robinson Dissimilarities
Mikhaël Carmona, Victor Chepoi, Guyslain Naves, and Pascal Préa

225 Hierarchies and Weak-hierarchies as Interval Convexities
Patrice Bertrand and Jean Diatta

12:00–13:00 **Room: 218**

CS **Contributed Session: Interpretable Machine Learning** [IML]

Chair: Johannes Fürnkranz

190 A Rule-based Approach to Scoring Systems
Michael Rapp, Johannes Fürnkranz, and Eyke Hüllermeier

193 On Explaining Model Change Based on Feature Importance
Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, and Eyke Hüllermeier

250 Interpretable Multi-class Trees for Travel Choice Mode Analysis
Christian Riccio, Andrea Papola, Michele Staiano, and Roberta Siciliano

Friday, 22 July

12:00–13:00 **Room: 215**

CS **Contributed Session: Classification Applied to Biological Sciences** [SPE]

Chair: Filipe Marques and Regina Bispo

128 Identification of Driver Genes in Glioblastoma via Regularized Classification
Marta Belchior Lopes and Susana Vinga

130 Outlier Detection: A Procedure to Capture Atypical Groups of Observations
Ana Helena Tavares, Vera Afreixo, and Paula Brito

211 Bayesian Classification and Non-Bayesian Label Estimation via EM Algorithm to Identify Differential Expression in Omics Data: a Comparative Study
Marília Antunes and Lisete Sousa

13:00–14:30 **Lunch Break**

14:30–15:30 **Room: Grand Hall**

KL **Keynote Lecture: Genevera Allen** [K3]

Chair: Rebecca Nugent

271 Fast Minipatch Ensemble Strategies for Learning and Inference
Genevera Allen

Friday, 22 July

15:30–16:00 **Room: Main Hall, Ground Floor**

PS **Poster Session 2** [PS2]

68 Parameter Estimation for Mixtures of Linear Mixed Models: the EM, CEM and SEM Algorithms
Luísa Novais and Susana Faria

101 Sequence-aware Item Recommendations for Multiply Repeated User-item Interactions
Juan Pablo Equihua, Maged Ali, Henrik Nordmark, and Berthold Lausen

194 High-dimensional Linear Regression Estimation
Mauro Iannuzzi and Matteo Farnè

213 Experimental Study of Similarity Measures for Clustering Uncertain Time Series
Michael Dinzinger, Michael Franklin Mbouopda, and Engelbert Mephu Nguifo

215 Assessing the Status of Two Data-Limited Skates Landed in Portuguese Ports Using an Empirical Catch Rule
Erick Chatalov, Ivone Figueiredo, Lisete Sousa, and Bárbara Pereira

216 Machine Learning Approach to Identify Factors that Influence Accident Severity
Daniel Santos, Vitor Nogueira, José Saias, Paulo Quaresma, Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Pedro Nogueira, Marcelo Silva, Rosalina Costa, Patrícia Gois, and Paulo Manuel

233 Hausdorff Distance: a Powerful Tool for Matching Households and Individuals in Historical Censuses
Thais Pacheco Menezes, Michael Fop, and Thomas Brendan Murphy

234 Model-based Tri-clustering
Rafika Boutalbi, Lazher Labiod, and Mohamed Nadif

244 Analyzing the Effects of Deviations from Normality on the Latent Growth Curve Models' Goodness-of-fit
Catarina Marques, Maria de Fátima Salgueiro, and Paula Cristina Vicente

249 Transformation Mixture Modeling for Skewed Data Groups with Heavy Tails
Yana Melnykov, Xuwen Zhu, and Volodymyr Melnykov

252 A Simulation Study on Variable Selection in Mixture Regression Models
Susana Faria

269 Hybrid Forecasting Combinations by Feature Based Metalearning
Moises Santos, André Carvalho, and Carlos Soares

277 Socio-Economic Classification of Territorial Units: Extreme Value Theory-based Methods as Support for the Construction of a Synthetic Index
Aleksandra Łuczak and Małgorzata Just

278 On the Measurement of Household Subjective Poverty: Concepts and Application
Aleksandra Łuczak and Sławomir Kalinowski

16:00–16:30 **Coffee Break**

Friday, 22 July

16:30–18:10 **Room: 113**

TT **Thematic Track: Functional Data Analysis 6** [FDA6]

Chair: Antonio Elias Fernandez

65 Depth-based Two-sample Testing
Felix Gnettner, Claudia Kirch, and Alicia Nieto-Reyes

169 The Control of False Discovery Rate for Functional Data
Niel Lundtorp Olsen, Alessia Pini, and Simone Vantini

202 Functional Random Forest for Biomedical Signals Classification and Interpretative Tools
Fabrizio Maturo and Rosanna Verde

253 Correlation-based Iterative Clustering Methods for Time Course Data
Michelle Carey, Shuang Wu, Guojun Gan, and Hulin Wu

280 Depth-based Classifiers for Partially Observed Functional Data
Antonio Elías, Raúl Jiménez, Anna M. Paganoni, and Laura M. Sangalli

16:30–17:50 **Room: 260**

TT **Thematic Track: Data Science for Business and Marketing** [DS-BM]

Chair: Andrzej Sokolowski

39 Using Clustering and Machine Learning Methods to Provide Intelligent Grocery Shopping Recommendations
Nail Chabane, Mohamed Achraf Bouaoune, Reda Amir Sofiane Tighilt, Bogdan Mazoure, Nadia Tahiri, and Vladimir Makarenkov

51 Typology of Motivation Factors for Employees in the Banking Sector: an Empirical Study Using Multivariate Data Analysis Methods
Áurea Sousa, Osvaldo Silva, M. Graça Batista, Sara Cabral, and Helena Bacelar-Nicolau

159 Industry Sector Detection in Legal Articles Using Transformer-based Deep Learning
Hui Yang, Stella Hadjiantoni, Yunfei Long, Ruta Petraityte, and Berthold Lausen

163 User Segmentation Based on Online Behavioural Data via Ensemble Predictions and Clustering
Stella Hadjiantoni, Hui Yang, Yunfei Long, Ruta Petraityte, and Berthold Lausen

16:30–17:50 **Room: 215**

TT **Thematic Track: Data Science in Education** [DS-E]

Chair: Pedro Campos

156 Attitudes Toward Statistics in the 3rd Cycle of Basic Education in Portugal
Adelaide Freitas, Ana Julieta Morais, Pedro Sá-Couto, and Anabela Rocha

241 Predictors of Quantitative Skills in Degree Schemes at University
Alex Partner, Adi Lausen, Alexei Vernitski, Chris Saker, and Berthold Lausen

247 Using Excel and R Teaching Statistics and Data Analysis
Wessel Moolman

248 Students' Assessment Through a IRT and Archetypal Analysis Joint Strategy
Lucio Palazzo and Francesco Palumbo

Friday, 22 July

16:30–17:50 **Room: 211**

TT **Thematic Track: Statistical Learning and Data Mining 3** [SLDM3]

Chair: Hyunjoong Kim

72 Kernel-based Hierarchical Structural Component Models for Pathway Analysis
Suhyun Hwangbo, Sungyoung Lee, Seungyeoun Lee, Heungsun Hwang, Inyoung Kim, and Taesung Park

87 Bayesian Inference for the Generation Interval of COVID-19 in Busan, Korea
Jayeong Paek, Ilsu Choi, Kyeongah Nah, and Yongkuk Kim

89 Fitting an Accelerated Failure Time Model with Time-dependent Covariates via Nonparametric Gaussian Scale Mixtures
Ju-Young Park, Sangwook Kang, and Byungtae Seo

146 Comparison of Survival Prediction Models for Pancreatic Cancer: Cox Model vs. Machine Learning Models
Hyunsuk Kim, Taesung Park, and Seungyeoun Lee

16:30–17:30 **Room: 118**

TT **Thematic Track: Model-based Clustering 6** [MBC6]

Chair: Volodymyr Melnykov

8 Clustering High-dimensional Microbiome Data
Sanjeena Dang (Subedi) and Wangshu Tu

12 Clustering Adolescent Female Physical Activity Levels with an Infinite Mixture Model on Random Effects
Amy LaLonde, Tanzy Love, Deborah Rohm Young, and Tongtong Wu

110 Modeling Three-way RNA Sequencing Data Using Data Transformations and Matrix-variate Gaussian Mixture Models
Theresa Scharl and Bettina Gruen

16:30–17:50 **Room: 159**

TT **Thematic Track: Robust Methods 4** [RM4]

Chair: Agustín Mayo Íscar

42 Some Issues in Robust Clustering
Christian Hennig

116 Robustness and Initialization Issues in Subspace Clustering
Luis Ángel García-Escudero and Agustín Mayo-Íscar

223 A Likelihood Ratio Test for Choosing Input Parameters in Robust Model Based Clustering
Luis Ángel García-Escudero, Agustín Mayo-Íscar, Gianluca Morelli, and Marco Riani

94 Assessing Common Principal Directions
David Rodríguez Vítores and Carlos Matrán Bea

Friday, 22 July

16:30–17:50 **Room: 213**

TT **Thematic Track: Social Network Analysis 2** [SNA2]

Chair: Giuseppe Giordano

36 Data Clustering and Representation Learning Based on Networked Data
Lazhar Labiod and Mohamed Nadif

224 Exploratory Graph Analysis for Configural Invariance Assessment of a Test
Alex Cucco, Lara Fontanella, Sara Fontanella, and Nicola Pronello

243 An Extension of Edge Reduction for Large Networks
Pedro Campos

245 Patterns of Cooperation for Polish Authors of Research Publications in Economics, Business, and Medicine Area
Paweł Lula, Urszula Cieraszewska, Magdalena Talaga, and Marcela Zembura

16:30–17:30 **Room: 135**

CS **Contributed Session: Clustering 6** [CL6]

Chair: Javier Trejos

16 A Topological Clustering of Individuals
Rafik Abdesselam

17 Modeling a Most Specific Generalization in Domain Taxonomies
Zhirayr Hayrapetyan, Boris Mirkin, Susana Nascimento, Trevor Fenner, and Dmitry Frolov

23 A Proposal for Formalization and Definition of Anomalies in Dynamical Systems
Jan Michael Spoor, Jens Weber, and Jivka Ovtcharova

16:30–17:50 **Room: 218**

CS **Contributed Session: Time Series Classification** [TSC]

Chair: Pablo Montero Manso

1 Unsupervised Classification of Categorical Time Series Through Innovative Distances
Ángel López-Oriona, José Antonio Vilar, and Pierpaolo D'Urso

22 Detecting Differences in Italian Regional Health Services During Two Covid-19 Waves
Lucio Palazzo and Riccardo Ievoli

118 The Clustering Performance of a Weighted Combined Distance Between Time Series
Margarida G. M. S. Cardoso, Ana Alexandra Martins, and João Lagarto

227 Dimension Reduction and Multivariate Time Series Classification
Veronne Yepmo, Angeline Plaud, and Engelbert Mephu Nguifo

18:00–19:30 **Room: Council Room**

CM **IFCS Council Meeting**

20:00 **Conference Diner** *Taylor's Port Wine Cellars*

Saturday, 23 July

8:30–9:00 **Registration**

9:00–10:30 **Room: 118**

IS **Invited Session: Dimension Reduction** [INV4]

Chair: Cinzia Viroli

200 Variable Screening in High Dimensional Regression via Random Projection Ensembles
Laura Anderlucci, Matteo Farnè, Giuliano Galimberti, and Angela Montanari

203 Model-based Clustering and Dimension Reduction for Multidimensional Social Networks
Michael Fop, Silvia D'Angelo, and Marco Alfò

220 Conditional Gaussian Mixture Modeling
Volodymyr Melnykov and Yang Wang

9:00–10:30 **Room: 218**

IS **Invited Session: Explainable Machine Learning** [INV5]

Chair: Eyke Hüllermeyer

259 Classification Over Text, Relational Databases and Graphs - Software and Case Studies
Tomáš Kliegr

274 Towards Deep and Interpretable Rule Learning
Johannes Fürnkranz

279 Current Challenges in Interpretable Machine Learning and Partitioning Approaches
Bernd Bischl

10:30–11:00 **Coffee Break**

11:00–12:00 **Room: 159**

TT **Thematic Track: Data Science in Economics and Finance 2** [DS-EF2]

Chair: Andrzej Dudek

35 A Review on Official Survey Item Classification for Mixed-Mode Effects Adjustment
Afshin Ashofteh and Pedro Campos

162 Adaptive Fuzzy Systems in Economics and Finance: Evaluating Interval Forecasts of High-Frequency Data
Rosangela Ballini

197 The Usefulness of Selected Machine Learning Methods for Estimating Missing Data to Supplement Databases Used for Corporate Bankruptcy Prediction
Barbara Pawełek and Józef Pociecha

Saturday, 23 July

11:00–12:00 **Room: 113**

TT **Thematic Track: Functional Data Analysis 7** [FDA7]

Chair: Eleonora Arnone

92 Registration of 24-hour Accelerometric Rest-activity Profiles and Its Application to Human Chronotypes
Erin McDonnell, Vadim Zipunnikov, Jennifer Schrack, Jeff Goldsmith, and Julia Wrobel

124 Functional Data from Wearable Devices. A Review
Nihan Acar-Denizli and Pedro Delicado

263 A Wavelet-mixed Effect Landmark Model for the Effect of Potassium and Biomarkers Profiles on Survival in Heart Failure Patients
Caterina Gregorio, Giulia Barbati, and Francesca Ieva

11:00–12:00 **Room: 211**

TT **Thematic Track: Optimization in Classification and Clustering 3** [OCC3]

Chair: Veronica Picciali

182 True Sparsity Approaches in Classification via Conic Optimization
Immanuel Bomze and Bo Peng

199 Creating Homogeneous Sectors: Criteria and Applications of Sectorization
Cristina Lopes, Maria Margarida Lima, Elif Goksu Ozturk, Ana Maria Rodrigues, Ana Catarina Nunes, Cristina Teles Oliveira, José Soeiro Ferreira, and Pedro Filipe Rocha

276 MARGOT: A Maximum MARGin Optimal Classification Tree
Federico D'Onofrio, Giorgio Grani, Marta Monaci, and Laura Palagi

11:00–12:00 **Room: 213**

TT **Thematic Track: Spatial-Temporal Data Analysis 2** [STDA2]

Chair: Koji Kurihara

69 Multivariate Mapping of Soil Organic Carbon and Nitrogen
Stephan van der Westhuizen, David Hofmeyr, and Gerard Heuvelink

120 Spatial Configuration of Fire Stations in Portugal
Regina Bispo, Clara Yokochi, Francisca Vieira, Nádia Bachir, Pedro Espadinha da Cruz, Pedro Lopes, Alexandre Penha, Marta Lopes, Filipe Marques, João Paulo Rodrigues, and António Grilo

142 Spatio-temporal Variability of Distribution and Abundance of Sardine in Portuguese Continental Coast: Environmental Effects
Daniela Silva, Raquel Menezes, Ana Moreno, Ana Teles-Machado, and Susana Garrido

Saturday, 23 July

11:00–12:00 **Room: 215**

CS **Contributed Session: Data Science in Biology** [DS-B]

Chair: Lisete Sousa

111 Time Resolved Feature Importance of a Biopharmaceutical Purification Process Using Permutation Based Methods
Matthias Medl, Theresa Scharl, Astrid Dürauer, and Friedrich Leisch

161 Off-Target Predictions in CRISPR-Cas9 Gene Editing Using Machine Learning
Ali Mertcan Kose and Ozan Kocadagli

238 Comparison of k-mer and Alignment-based Pre-processing Approaches for Machine Learning Based Functional Annotation with 16S rRNA Data
Rafal Kulakowski, Adi Lausen, Etienne Low-Decarie, and Berthold Lausen

11:00–12:00 **Room: 135**

CS **Contributed Session: Dimension Reduction & Clustering** [DR&CL]

Chair: Francesco Palumbo

63 An Ultrametric Model for Clustering and Dimensionality Reduction
Giorgia Zaccaria

66 Combining Latent Class Analysis and Multiple Correspondence Analysis
Alice Barth

166 Simultaneous Factorial Reduction and Clustering for Three-mode Data Sets: a Comparison
Prosper Ablordeppey, Adelaide Freitas, Maurizio Vichi, and Giorgia Zaccaria

11:00–12:00 **Room: 118**

CS **Contributed Session: Model-based Clustering for Time Series** [MBC-TS]

Chair: Laura Anderlucci

70 Clustering Intensive Longitudinal Data Through Mixture Multilevel Vector-autoregressive Modeling
Anja Ernst, Marieke Timmerman, Feng Ji, Bertus Jeronimus, and Casper Albers

136 Mispecification Tests for Hidden Markov Models Based on a New Class of Finite Mixture Models
Francesco Bartolucci, Silvia Pandolfi, and Fulvia Pennoni

179 Natural Cubic Smoothing Splines for Latent Class Identification in Longitudinal Growth Trajectories
Katerina M. Marcoulides and Laura Trinchera

Saturday, 23 July

11:00–12:00 **Room: 218**

CS **Contributed Session: Machine Learning 4** [ML4]

Chair: Bernd Bischl

4 Stochastic Collapsed Variational Inference for Structured Gaussian Process Regression Networks
Rui Meng, Herbert Lee, and Kristofer Bouchard

191 Reliability Assessment of Ancient Stone Arch Bridge Applying ANN Models, Case Study: Leça Railway Bridge
Edward Baron, Ana Margarida Bento, José Campos e Matos, Rui Calçada, and Kenneth Gavin

196 Application of Artificial Intelligence (AI) in Flood Risk Forecasting
Minh Quang Tran, Ana Margarida Bento, Elisabete Teixeira, Hélder Sousa, and José Campos e Matos

11:00–12:00 **Room: 260**

CS **Contributed Session: Supervised Classification 2** [SC2]

Chair: Pedro Duarte Silva

104 Logistic Regression with Sparse Common and Distinctive Covariates
Soogeun Park, Eva Ceulemans, and Katrijn Van Deun

115 Accuracy Measures for Binary Classification Based on Quantitative Group Tests
Rui Santos, João Paulo Martins, and Miguel Felgueiras

175 Exploiting Pareto Density Estimation for Nonparametric Naive Bayes Classifiers
Quirin Stier and Michael Thrun

12:00–13:00 **Room: Grand Hall**

KL **Keynote Lecture: João Gama** [K4]

Chair: Berthold Lausen

62 Trends in Data Stream Mining
João Gama

13:00–13:30 **Room: Grand Hall**

CLS **Closing Session**

13:30–14:30 **Lunch**