

Irena Koprinska · Paolo Mignone ·
Riccardo Guidotti · Szymon Jaroszewicz ·
Holger Fröning · Francesco Gullo ·
Pedro M. Ferreira · Damian Roqueiro et al. (Eds.)

Communications in Computer and Information Science

1753

Machine Learning and Principles and Practice of Knowledge Discovery in Databases

International Workshops of ECML PKDD 2022
Grenoble, France, September 19–23, 2022
Proceedings, Part II

Part 2

 Springer

Editorial Board Members

Joaquim Filipe 

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Raquel Oliveira Prates 

Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <https://link.springer.com/bookseries/7899>

Irena Koprinska · Paolo Mignone ·
Riccardo Guidotti · Szymon Jaroszewicz ·
Holger Fröning · Francesco Gullo ·
Pedro M. Ferreira · Damian Roqueiro et al. (Eds.)

Machine Learning and Principles and Practice of Knowledge Discovery in Databases

International Workshops of ECML PKDD 2022
Grenoble, France, September 19–23, 2022
Proceedings, Part II

For the full list of editors *see next page*

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-23632-7 ISBN 978-3-031-23633-4 (eBook)
<https://doi.org/10.1007/978-3-031-23633-4>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023

Chapter “On the Development of a European Tracker of Societal Issues and Economic Activities Using Alternative Data” is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). For further details see license information in the chapter.

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.


The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Editors

Irena Koprinska 
University of Sydney
Sydney, Australia

Riccardo Guidotti 
University of Pisa
Pisa, Italy

Holger Fröning 
Heidelberg University
Heidelberg, Germany

Pedro M. Ferreira 
University of Lisbon
Lisbon, Portugal

Gaia Ceddia
Barcelona Supercomputing Center
Barcelona, Spain

João Gama 
University of Porto
Porto, Portugal

Ricard Gavaldà 
UPC BarcelonaTech
Barcelona, Spain

Zbigniew Ras 
University of North Carolina
Charlotte, USA

Francesca Naretto
University of Pisa
Pisa, Italy

Przemyslaw Biecek 
Warsaw University of Technology
Warsaw, Poland

Gregor Schiele 
University of Duisburg-Essen
Essen, Germany

Michaela Blott
AMD
Dublin, Ireland


Ivan Luciano Danesi
UniCredit
Milan, Italy


Paolo Mignone
University of Bari Aldo Moro
Bari, Italy

Szymon Jaroszewicz
Warsaw University of Technology
Warsaw, Poland

Francesco Gullo
UniCredit
Rome, Italy

Damian Roqueiro
Roche
Basel, Switzerland

Sławomir Nowaczyk 
Halmstad University
Halmstad, Sweden

Rita Ribeiro 
University of Porto
Porto, Portugal

Elio Masciari 
University of Naples Federico II
Naples, Italy

Ettore Ritacco
ICAR-CNR
Rende, Italy

Andreas Theissler
Aalen University of Applied Sciences
Aalen, Germany


Wouter Verbeke
KU Leuven
Leuven, Belgium


Franz Pernkopf 
Graz University of Technology
Graz, Austria


Ilaria Bordino
UniCredit
Rome, Italy

Giovanni Ponti
National Agency for New Technologies
Rome, Italy

Lorenzo Severini
Unicredit
Rome, Italy

Giuseppina Andresini 
University of Bari Aldo Moro
Bari, Italy


Guilherme Graça 
University of Lisbon
Lisbon, Portugal

Naghmeh Ghazaleh 
Roche
Basel, Switzerland

Diego Saldana
Novartis
Basel, Switzerland

Arif Canakoglu
Fondazione IRCCS Ca' Granda Ospedale
Maggiore Policlinico
Milan, Italy


Pietro Pinoli
Politecnico di Milano
Milan, Italy


Sepideh Pashami 
Halmstad University
Halmstad, Sweden

Annalisa Appice 
University of Bari Aldo Moro
Bari, Italy

Ibéria Medeiros
University of Lisbon
Lisbon, Portugal

Lee Cooper 
Northwestern University
Chicago, USA

Jonas Richiardi 
University of Lausanne
Lausanne, Switzerland

Konstantinos Sechidis 
Novartis
Basel, Switzerland

Sara Pido
Politecnico di Milano
Milan, Italy

Albert Bifet 
University of Waikato
Hamilton, New Zealand

Preface

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) is the premier European conference on machine learning and data mining. In 2022, ECML PKDD took place in Grenoble, France during September 19–23.

The program included workshops on specialized topics held during the first and last day of the conference. This two-volume set includes the proceedings of the following workshops:

1. 7th Workshop on Data Science for Social Good (SoGood 2022)
2. 10th Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2022)
3. 4th Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2022)
4. 1st Workshop on Uplift Modeling (UMOD 2022)
5. 3rd Workshop on IoT, Edge and Mobile for Embedded Machine Learning (ITEM 2022)
6. 7th Workshop on Mining Data for Financial Application (MIDAS 2022)
7. 4th Workshop on Machine Learning for Cybersecurity (MLCS 2022)
8. 2nd Workshop on Machine Learning for Buildings Energy Management (MLBEM 2022)
9. 3rd Workshop on Machine Learning for Pharma and Healthcare Applications (PharML 2022)
10. 1st Workshop on Data Analysis in Life Science (DALIS 2022)
11. 3rd Workshop on IoT Streams for Predictive Maintenance (IoT-PdM 2022)

Each workshop section contains the papers from the workshop and a preface from the organizers.

We would like to thank all participants and invited speakers, the Program Committees and reviewers, and the ECML PKDD conference and workshop chairs – thank you for making the workshops successful events. We are also grateful to Springer for their help in publishing the proceedings.

October 2022

Irena Koprinska
on behalf of the volume editors

Organization

ECML PKDD 2022 Workshop Chairs

Bruno Crémilleux
Charlotte Laclau

Université de Caen Normandie, France
Télécom Paris, France

SoGood 2022 Chairs

João Gama
Irena Koprinska
Rita P. Ribeiro
Ricard Gavaldà

University of Porto, Portugal
University of Sydney, Australia
University of Porto, Portugal
BarcelonaTech, Spain

NFMCP 2022 Chairs

Elio Masciari
Paolo Mignone
Zbigniew W. Ras
Ettore Ritacco

University Federico II of Naples, Italy
University of Bari Aldo Moro, Italy
University of North Carolina, USA
ICAR-CNR, Italy

XKDD 2022 Chairs

Riccardo Guidotti
Francesca Naretto
Andreas Theissler
Przemysław Biecek

University of Pisa, Italy
Scuola Normale Superiore, Italy
Aalen University of Applied Sciences, Germany
Warsaw University of Technology, Poland

UMOD 2022 Chairs

Szymon Jaroszewicz
Wouter Verbeke

Polish Academy of Sciences and Warsaw
University of Technology, Poland
KU Leuven, Belgium

ITEM 2022 Chairs

Holger Fröning
Gregor Schiele
Franz Pernkopf
Michaela Blott

Heidelberg University, Germany
University of Duisburg-Essen, Germany
Graz University of Technology, Austria
AMD, Dublin, Ireland

MIDAS 2022 Chairs

Ilaria Bordino	UniCredit, Italy
Ivan Luciano Danesi	UniCredit, Italy
Francesco Gullo	UniCredit, Italy
Giovanni Ponti	ENEA, Italy
Lorenzo Severini	UniCredit, Italy

MLCS 2022 Chairs

Pedro M. Ferreira	University of Lisbon, Portugal
Annalisa Appice	University of Bari, Italy
Giuseppina Andresini	University of Bari, Italy
Ibéria Medeiros	University of Lisbon, Portugal

MLBEM 2022 Chairs

Pedro M. Ferreira	University of Lisbon, Portugal
Guilherme Graça	University of Lisbon, Portugal

PharML 2022 Chairs

Damian Roqueiro	Roche, Basel, Switzerland
Lee Cooper	Northwestern University, USA
Naghme Ghazaleh	Roche, Basel, Switzerland
Jonas Richiardi	Lausanne University Hospital and University of Lausanne, Switzerland
Diego Saldana	Novartis, Basel, Switzerland
Konstantinos Sechidis	Novartis, Basel, Switzerland

DALS 2022 Chairs

Gaia Ceddia	Barcelona Supercomputing Center, Spain
Arif Canakoglu	Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
Sara Pido	Massachusetts Institute of Technology, USA, and Politecnico di Milano, Italy
Pietro Pinoli	Politecnico di Milano, Italy

IoT-PdM 2022 Chairs

Joao Gama	University of Porto, Portugal
Albert Bifet	Telecom-Paris, Paris, France, and University of Waikato, New Zealand
Sławomir Nowaczyk	Halmstad University, Sweden
Sepideh Pashami	Halmstad University, Sweden

Contents – Part II

Workshop on Mining Data for Financial Application (MIDAS 2022)

Multi-task Learning for Features Extraction in Financial Annual Reports	7
<i>Syrielle Montariol, Matej Martinc, Andraž Pelicon, Senja Pollak, Boshko Koloski, Igor Lončarski, Aljoša Valentinčič, Katarina Sitar Šuštar, Riste Ichev, and Martin Žnidaršič</i>	
What to Do with Your Sentiments in Finance	25
<i>Argimiro Arratia</i>	
On the Development of a European Tracker of Societal Issues and Economic Activities Using Alternative Data	38
<i>Sergio Consoli, Marco Colagrossi, Francesco Panella, and Luca Barbaglia</i>	
Privacy-Preserving Machine Learning in Life Insurance Risk Prediction	44
<i>Klismam Pereira, João Vinagre, Ana Nunes Alonso, Fábio Coelho, and Melânia Carvalho</i>	
Financial Distress Model Prediction Using Machine Learning: A Case Study on Indonesia’s Consumers Cyclical Companies	53
<i>Niken Prasasti Martono and Hayato Ohwada</i>	
Improve Default Prediction in Highly Unbalanced Context	62
<i>Stefano Piersanti</i>	
Towards Explainable Occupational Fraud Detection	79
<i>Julian Tritscher, Daniel Schlör, Fabian Gwinner, Anna Krause, and Andreas Hotho</i>	
Towards Data-Driven Volatility Modeling with Variational Autoencoders	97
<i>Thomas Dierckx, Jesse Davis, and Wim Schoutens</i>	
Auto-clustering of Financial Reports Based on Formatting Style and Author’s Fingerprint	112
<i>Braulio C. Blanco Lambruschini, Mats Brorsson, and Maciej Zurad</i>	

InFi-BERT 1.0: Transformer-Based Language Model for Indian Financial Volatility Prediction 128
Sravani Sasubilli and Mridula Verma

Workshop on Machine Learning for Cybersecurity (MLCS 2022)

Intrusion Detection Using Ensemble Models 143
Tina Yazdizadeh, Shabnam Hassani, and Paula Branco

Domain Adaptation with Maximum Margin Criterion with Application to Network Traffic Classification 159
Zahra Taghiyarrenani and Hamed Farsi

Evaluation of the Limit of Detection in Network Dataset Quality Assessment with PerQoDA 170
Katarzyna Wasielewska, Dominik Soukup, Tomáš Čejka, and José Camacho

Towards a General Model for Intrusion Detection: An Exploratory Study 186
Tommaso Zoppi, Andrea Ceccarelli, and Andrea Bondavalli

Workshop on Machine Learning for Buildings Energy Management (MLBEM 2022)

Conv-NILM-Net, a Causal and Multi-appliance Model for Energy Source Separation 207
Mohamed Alami C., Jérémie Decock, Rim kaddah, and Jesse Read

Domestic Hot Water Forecasting for Individual Housing with Deep Learning 223
Paul Compagnon, Aurore Lomet, Marina Reyboz, and Martial Mermillod

Workshop on Machine Learning for Pharma and Healthcare Applications (PharML 2022)

Detecting Drift in Healthcare AI Models Based on Data Availability 243
Ylenia Rotalinti, Allan Tucker, Michael Lonergan, Puja Myles, and Richard Branson

Assessing Different Feature Selection Methods Applied to a Bulk RNA Sequencing Dataset with Regard to Biomedical Relevance 259
Damir Zhakparov, Kathleen Moriarty, Nonhlanhla Lunjani, Marco Schmid, Carol Hlela, Michael Levin, Avumile Mankahla, SOS-ALL Consortium, Cezmi Akdis, Liam O’Mahony, Katja Baerenfaller, and Damian Roqueiro

Predicting Drug Treatment for Hospitalized Patients with Heart Failure	275
<i>Linyi Zhou and Ioanna Miliou</i>	
A Workflow for Generating Patient Counterfactuals in Lung Transplant Recipients	291
<i>Franco Rugolon, Maria Bampa, and Panagiotis Papapetrou</i>	
Few-Shot Learning for Identification of COVID-19 Symptoms Using Generative Pre-trained Transformer Language Models	307
<i>Keyuan Jiang, Minghao Zhu, and Gordon R. Bernard</i>	
A Light-Weight Deep Residual Network for Classification of Abnormal Heart Rhythms on Tiny Devices	317
<i>Rohan Banerjee and Avik Ghose</i>	
Workshop on Data Analysis in Life Science (DALIS 2022)	
I-CONVEX: Fast and Accurate <i>de Novo</i> Transcriptome Recovery from Long Reads	339
<i>Sina Baharlouei, Meisam Razaviyayn, Elizabeth Tseng, and David Tse</i>	
Italian Debate on Measles Vaccination: How Twitter Data Highlight Communities and Polarity	364
<i>Cynthia Ifeyinwa Ugwu and Sofia Casarin</i>	
3rd Workshop and Tutorial on Streams for Predictive Maintenance (IoT-PdM 2022)	
Online Anomaly Explanation: A Case Study on Predictive Maintenance	383
<i>Rita P. Ribeiro, Saulo Martiello Mastelini, Narjes Davari, Ehsan Aminian, Bruno Veloso, and João Gama</i>	
Fault Forecasting Using Data-Driven Modeling: A Case Study for Metro do Porto Data Set	400
<i>Narjes Davari, Bruno Veloso, Rita P. Ribeiro, and João Gama</i>	
An Online Data-Driven Predictive Maintenance Approach for Railway Switches	410
<i>Emanuel Sousa Tomé, Rita P. Ribeiro, Bruno Veloso, and João Gama</i>	
curr2vib: Modality Embedding Translation for Broken-Rotor Bar Detection	423
<i>Amirhossein Berenji, Zahra Taghiyarrenani, and Sławomir Nowaczyk</i>	

Incorporating Physics-Based Models into Data Driven Approaches for Air Leak Detection in City Buses	438
<i>Yuantao Fan, Hamid Sarmadi, and Sławomir Nowaczyk</i>	
Towards Geometry-Preserving Domain Adaptation for Fault Identification	451
<i>Zahra Taghiyarrenani, Sławomir Nowaczyk, Sepideh Pashami, and Mohamed-Rafik Bouguelia</i>	
A Systematic Approach for Tracking the Evolution of XAI as a Field of Research	461
<i>Samaneh Jamshidi, Sławomir Nowaczyk, Hadi Fanaee-T, and Mahmoud Rahat</i>	
Frequent Generalized Subgraph Mining via Graph Edit Distances	477
<i>Richard Palme and Pascal Welke</i>	
Author Index	485