

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 <http://www.springer.com/series/7899>

Irena Koprinska · Michael Kamp ·  
Annalisa Appice · Corrado Loglisci ·  
Luiza Antonie · Albrecht Zimmermann ·  
Riccardo Guidotti · Özlem Özgöbek et al. (Eds.)

# ECML PKDD 2020 Workshops

Workshops of the European Conference on Machine  
Learning and Knowledge Discovery in Databases  
(ECML PKDD 2020): SoGood 2020, PDFL 2020,  
MLCS 2020, NFMCP 2020, DINA 2020,  
EDML 2020, XKDD 2020 and INRA 2020  
Ghent, Belgium, September 14–18, 2020  
Proceedings

For the full list of editors *see next page*

ISSN 1865-0929                      ISSN 1865-0937 (electronic)  
Communications in Computer and Information Science  
ISBN 978-3-030-65964-6              ISBN 978-3-030-65965-3 (eBook)  
<https://doi.org/10.1007/978-3-030-65965-3>

© Springer Nature Switzerland AG 2020

7 chapters are licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). For further details see licence information in the chapters.

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 Koprinksa 

University of Sydney  
Sydney, NSW, Australia

Annalisa Appice 

University of Bari Aldo Moro  
Bari, Italy

Luiza Antonie 

University of Guelph  
Guelph, ON, Canada

Riccardo Guidotti 


University of Pisa  
Pisa, Italy

Rita P. Ribeiro 

University of Porto  
Porto, Portugal

João Gama 

University of Porto  
Porto, Portugal

Yamuna Krishnamurthy 

Royal Holloway University of London  
Egham, UK

Donato Malerba 

University of Bari Aldo Moro  
Bari, Italy

Michelangelo Ceci 

University of Bari Aldo Moro  
Bari, Italy

Elio Masciari 

University of Naples Federico II  
Naples, Italy

Peter Christen 

Australian National University  
Canberra, ACT, Australia

Erich Schubert

Technical University of Dortmund  
Dortmund, Germany

Anna Monreale 

University of Pisa  
Pisa, Italy

Salvatore Rinzivillo 

ISTI-CNR  
PISA, Italy

Andreas Lommatzsch

Berlin Institute of Technology  
Berlin, Germany

Michael Kamp 

Monash University  
Clayton, VIC, Australia

Corrado Loglisci 

University of Bari Aldo Moro  
Bari, Italy

Albrecht Zimmermann 

University of Caen Normandy  
Caen, France

Özlem Özgöbek

Norwegian University of Science  
and Technology  
Trondheim, Norway

Ricard Gavalda 

UPC BarcelonaTech  
Barcelona, Spain

Linara Adilova

Fraunhofer IAIS  
St. Augustin, Germany

Pedro M. Ferreira 

University of Lisbon  
Lisbon, Portugal

Ibéria Medeiros 

University of Lisbon  
Lisbon, Portugal

Giuseppe Manco 

ICAR-CNR  
Rende, Italy

Zbigniew W. Ras 


University of North Carolina  
Charlotte, NC, USA

Eirini Ntouts


Leibniz University Hannover  
Hannover, Germany

Arthur Zimek 

University of Southern Denmark  
Odense, Denmark

Przemyslaw Biecek 

Warsaw University of Technology  
Warsaw, Poland

Benjamin Kille 

Berlin Institute of Technology  
Berlin, Germany

Jon Atle Gulla

Norwegian University of Science  
and Technology  
Trondheim, Norway

# Preface

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2020) is the premier European conference on machine learning and data mining. ECML PKDD was scheduled to take place in Ghent, Belgium, but due to the COVID-19 pandemic it was held fully virtually during September 14–18, 2020.

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

1. Fifth Workshop on Data Science for Social Good (SoGood 2020)
2. Workshop on Parallel, Distributed and Federated Learning (PDFL 2020)
3. Second Workshop on Machine Learning for Cybersecurity (MLCS 2020)
4. Ninth International Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2020)
5. Workshop on Data Integration and Applications (DINA 2020)
6. Second Workshop on Evaluation and Experimental Design in Data Mining and Machine Learning (EDML 2020)
7. Second International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2020)
8. Eighth International Workshop on News Recommendation and Analytics (INRA 2020)

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 a successful event. We are also grateful to Springer for their help in publishing this volume.

October 2020

Irena Koprinska  
on behalf of the volume editors

# Organization

## Workshop Chairs

### SoGood 2020

Ricard Gavalda  
Irena Koprinska  
João Gama  
Rita P. Ribeiro

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

### PDFL 2020

Michael Kamp  
Linara Adilova  
Yamuna Krishnamurthy

Monash University, Australia  
Fraunhofer IAIS, Germany  
Royal Holloway, University of London, UK

### MLCS 2020

Annalisa Appice  
Pedro M. Ferreira  
Michael Kamp  
Donato Malerba  
Ibéria Medeiros

University of Bari Aldo Moro, Italy  
University of Lisbon, Portugal  
Monash University, Australia  
University of Bari Aldo Moro, Italy  
University of Lisbon, Portugal

### NFMCP 2020

Michelangelo Ceci  
Corrado Loglisci  
Giuseppe Manco  
Elio Masciari  
Zbigniew W. Ras

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

### DINA 2020

Luiza Antonie  
Peter Christen  
Erhard Rahm  
Osmar Zaiane

University of Guelph, Canada  
The Australian National University, Australia  
University of Leipzig, Germany  
University of Alberta, Canada

## **EDML 2020**

Eirini Ntoutsis  
Erich Schubert  
Athur Zimek  
Albrecht Zimmermann

Leibniz University Hannover, Germany  
Technical University of Dortmund, Germany  
University of Southern Denmark, Denmark  
University of Caen Normandy, France

## **XKDD 2020**

Riccardo Guidotti  
Anna Monreale  
Salvatore Rinzivillo  
Przemyslaw Biecek

University of Pisa, Italy  
University of Pisa, Italy  
ISTI-CNR, Italy  
Warsaw University of Technology, Poland

## **INRA 2020**

Özlem Özgöbek  
Benjamin Kille  
Andreas Lommatzsch  
Jon Atle Gulla

Norwegian University of Science and Technology,  
Norway  
Berlin Institute of Technology, Germany  
Berlin Institute of Technology, Germany  
Norwegian University of Science and Technology,  
Norway



# Contents

## **Fifth Workshop on Data Science for Social Good (SoGood 2020)**

On Modeling Labor Markets for Fine-Grained Insights . . . . .	9
<i>Hendrik Santoso Sugiarto and Ee-Peng Lim</i>	
Reasoning About Neural Network Activations: An Application in Spatial Animal Behaviour from Camera Trap Classifications . . . . .	26
<i>Benjamin C. Evans, Allan Tucker, Oliver R. Wearn, and Chris Carbone</i>	
Practical Lessons from Generating Synthetic Healthcare Data with Bayesian Networks . . . . .	38
<i>Juan de Benedetti, Namir Oues, Zhenchen Wang, Puja Myles, and Allan Tucker</i>	
Building Trajectories Over Topology with TDA-PTS: An Application in Modelling Temporal Phenotypes of Disease . . . . .	48
<i>Seyed Erfan Sajjadi, Barbara Draghi, Lucia Sacchi, Arianna Dagliani, John Holmes, and Allan Tucker</i>	
Data Decomposition Based Learning for Load Time-Series Forecasting . . . . .	62
<i>Jatin Bedi and Durga Toshniwal</i>	
Mitigating Bias in Online Microfinance Platforms: A Case Study on Kiva.org . . . . .	75
<i>Soumajyoti Sarkar and Hamidreza Alvani</i>	
A Left Realist Critique of the Political Value of Adopting Machine Learning Systems in Criminal Justice . . . . .	92
<i>Fabio Massimo Zennaro</i>	

## **Workshop on Parallel, Distributed and Federated Learning (PDFL 2020)**

Knowledge Discovery on Blockchains: Challenges and Opportunities for Distributed Event Detection Under Constraints . . . . .	113
<i>Cedric Sanders and Thomas Liebig</i>	
Resource-Constrained On-Device Learning by Dynamic Averaging . . . . .	129
<i>Lukas Hepp, Michael Kamp, Linara Adilova, Danny Heinrich, Nico Piatkowski, and Katharina Morik</i>	

## **Second Workshop on Machine Learning for Cybersecurity (MLCS 2020)**

MitM Attack Detection in BLE Networks Using Reconstruction and Classification Machine Learning Techniques. . . . .	149
<i>Abdelkader Lahmadi, Alexis Duque, Nathan Heraief, and Julien Francq</i>	
Advocating for Multiple Defense Strategies Against Adversarial Examples. . .	165
<i>Alexandre Araujo, Laurent Meunier, Rafael Pinot, and Benjamin Negrevergne</i>	
Hybrid Connection and Host Clustering for Community Detection in Spatial-Temporal Network Data . . . . .	178
<i>Mark Patrick Roeling, Azqa Nadeem, and Sicco Verwer</i>	
Collaborative Learning Based Effective Malware Detection System. . . . .	205
<i>Narendra Singh, Harsh Kasyap, and Somanath Tripathy</i>	

## **Ninth International Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2020)**

A Hybrid Recommendation System Based on Bidirectional Encoder Representations . . . . .	225
<i>Irem Islek and Sule Gunduz Oguducu</i>	
Leveraging Multi-target Regression for Predicting the Next Parallel Activities in Event Logs. . . . .	237
<i>Michelangelo Ceci, Angelo Impedovo, and Antonio Pellicani</i>	
A Multi-view Ensemble of Deep Models for the Detection of Deviant Process Instances . . . . .	249
<i>Francesco Folino, Gianluigi Folino, Massimo Guarascio, and Luigi Pontieri</i>	
Exploiting Temporal Convolution for Activity Prediction in Process Analytics . . . . .	263
<i>Francesco Folino, Massimo Guarascio, Angelica Liguori, Giuseppe Manco, Luigi Pontieri, and Ettore Ritacco</i>	

## **Workshop on Data Integration and Applications (DINA 2020)**

Hyper-Parameter Optimization for Privacy-Preserving Record Linkage. . . . .	281
<i>Joyce Yu, Jakub Nabaglo, Dinusha Vatsalan, Wilko Henecka, and Brian Thorne</i>	
Group-Specific Training Data. . . . .	297
<i>Ben Busath, Jalen Morgan, and Joseph Price</i>	

Scalable Blocking for Very Large Databases . . . . .	303
<i>Andrew Borthwick, Stephen Ash, Bin Pang, Shehzad Qureshi, and Timothy Jones</i>	
Address Validation in Transportation and Logistics: A Machine Learning Based Entity Matching Approach . . . . .	320
<i>Yassine Guermazi, Sana Sellami, and Omar Boucelma</i>	
Linking Heterogeneous Data for Food Security Prediction . . . . .	335
<i>Hugo Deléglise, Agnès Bégué, Roberto Interdonato, Elodie Maître d'Hôtel, Mathieu Roche, and Maguelonne Teisseire</i>	
<b>Second Workshop on Evaluation and Experimental Design in Data Mining and Machine Learning (EDML 2020)</b>	
Towards Better Evaluation of Multi-target Regression Models . . . . .	353
<i>Evgeniya Korneva and Hendrik Blockeel</i>	
Assessing the Difficulty of Labelling an Instance in Crowdsourcing. . . . .	363
<i>Neetha Jambigi, Tirtha Chanda, Vishnu Unnikrishnan, and Myra Spiliopoulou</i>	
Experimental Evaluation of Scale, and Patterns of Systematic Inconsistencies in Google Trends Data. . . . .	374
<i>Philipp Behnen, Rene Kessler, Felix Kruse, Jorge Marx Gómez, Jan Schoenmakers, and Sergej Zerr</i>	
Assessing the Uncertainty of the Text Generating Process Using Topic Models . . . . .	385
<i>Jonas Rieger, Carsten Jentsch, and Jörg Rahnenführer</i>	
A Ranking Stability Measure for Quantifying the Robustness of Anomaly Detection Methods . . . . .	397
<i>Lorenzo Perini, Connor Galvin, and Vincent Vercruyssen</i>	
<b>Second International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2020)</b>	
Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges . . . . .	417
<i>Christoph Molnar, Giuseppe Casalicchio, and Bernd Bischl</i>	
Efficient Estimation of General Additive Neural Networks: A Case Study for CTG Data. . . . .	432
<i>P. J. G. Lisboa, S. Ortega-Martorell, M. Jayabalan, and I. Olier</i>	

What Would You Ask the Machine Learning Model? Identification of User Needs for Model Explanations Based on Human-Model Conversations . . . . .	447
<i>Michał Kuźba and Przemysław Biecek</i>	
Analyzing Forward Robustness of Feedforward Deep Neural Networks with LeakyReLU Activation Function Through Symbolic Propagation . . . . .	460
<i>Giulio Masetti and Felicita Di Giandomenico</i>	
LimeOut: An Ensemble Approach to Improve Process Fairness . . . . .	475
<i>Vaishnavi Bhargava, Miguel Couceiro, and Amedeo Napoli</i>	
Interpretable Privacy with Optimizable Utility. . . . .	492
<i>Jan Ramon and Moitree Basu</i>	
Prediction and Explanation of Privacy Risk on Mobility Data with Neural Networks . . . . .	501
<i>Francesca Naretto, Roberto Pellungrini, Franco Maria Nardini, and Fosca Giannotti</i>	
Approximate Explanations for Classification of Histopathology Patches . . . . .	517
<i>Iam Palatnik de Sousa, Marley M. B. R. Vellasco, and Eduardo Costa da Silva</i>	
<b>Eighth International Workshop on News Recommendation and Analytics (INRA 2020)</b>	
Multi-stakeholder News Recommendation Using Hypergraph Learning . . . . .	531
<i>Alireza Gharahighehi, Celine Vens, and Konstantinos Pliakos</i>	
Monitoring Technoscientific Issues in the News . . . . .	536
<i>Alberto Cammozzo, Emanuele Di Buccio, and Federico Neresini</i>	
Pitch Proposal: Recommenders with a Mission - Assessing Diversity in News Recommendations . . . . .	554
<i>Sanne Vrijenhoek and Natali Helberger</i>	
An Educational News Dataset for Recommender Systems . . . . .	562
<i>Yujie Xing, Itishree Mohallick, Jon Atle Gulla, Özlem Özgöbek, and Lemei Zhang</i>	
Exploring Thematic Coherence in Fake News. . . . .	571
<i>Martins Samuel Dogo, Deepak P., and Anna Jurek-Loughrey</i>	
Media Bias in German News Articles: A Combined Approach . . . . .	581
<i>Timo Spinde, Felix Hamborg, and Bela Gipp</i>	

On the Coherence of Fake News Articles. . . . .	591
<i>Iknoor Singh, Deepak P., and Anoop K.</i>	
<b>Short Paper – DINA 2020</b>	
Integrating Spreadsheets by Identifying and Solving Layout Constraints. . . . .	611
<i>Dirko Coetsee, McElory Hoffmann, and Luc De Raedt</i>	
<b>Author Index . . . . .</b>	<b>613</b>