

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>

Michael Kamp · Irena Koprinska · Adrien Bibal ·
Tassadit Bouadi · Benoît Frénay · Luis Galárraga ·
José Oramas · Linara Adilova et al. (Eds.)

Machine Learning and Principles and Practice of Knowledge Discovery in Databases

International Workshops of ECML PKDD 2021
Virtual Event, September 13–17, 2021
Proceedings, Part I

For the full list of editors *see next page*

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-93735-5 ISBN 978-3-030-93736-2 (eBook)
<https://doi.org/10.1007/978-3-030-93736-2>

© Springer Nature Switzerland AG 2021

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

Michael Kamp 
IKIM, Ruhr-University Bochum
Bochum, Germany

Monash University
Melbourne, Australia

Adrien Bibal 
University of Namur
Namur, Belgium

Benoît Frénay 
University of Namur
Namur, Belgium

José Oramas 
University of Antwerp
Antwerp, Belgium

Yamuna Krishnamurthy 
Royal Holloway University of London
Egham, UK

Christine Largeron
Université Jean Monnet
Saint-Etienne cedex 2, France

Tiphaine Viard
Telecom Paris
Paris, France

Massimiliano Ruocco
Norwegian Univesity of Science
and Technology
Trondheim, Norway

Claudio Gallicchio
University of Pisa
Pisa, Italy

Franz Pernkopf 
Graz University of Technology
Graz, Austria

Holger Fröning 
Heidelberg University
Heidelberg, Germany

Riccardo Guidotti 
University of Pisa
Pisa, Italy

Irena Koprinska 
University of Sydney
Sydney, NSW, Australia

Tassadit Bouadi 
University of Rennes 1
Rennes, France

Luis Galárraga 
Inria
Rennes, France

Linara Adilova
Ruhr University Bochum
Bochum, Germany

Bo Kang 
Ghent University
Ghent, Belgium

Jefrey Lijffijt 
Ghent University
Gent, Belgium

Pascal Welke 
University of Bonn
Bonn, Germany

Erlend Aune
BI Norwegian Business School
Oslo, Norway

Gregor Schiele 
University of Duisburg-Essen
Essen, Germany

Michaela Blott 
Xilinx Research
Dublin, Ireland

Günther Schindler
Heidelberg University
Heidelberg, Germany

Anna Monreale 
University of Pisa
Pisa, Italy

Salvatore Rinzivillo 
ISTI-CNR
Pisa, Italy

Eirini Ntoutsis 
Freie Universität Berlin
Berlin, Germany

Bodo Rosenhahn 
Leibniz University Hannover
Hannover, Germany

Daniela Cialfi 
University of Chieti-Pescara
Chieti, Italy

Maxwell Ramstead 
McGill University
Montreal, Canada

Pedro M. Ferreira 
University of Lisbon
Lisboa, Portugal

Donato Malerba 
Universita di Bari Aldo Moro
Bari, Italy

Philippe Fournier-Viger 
Shenzhen University
Shenzhen, China

Sebastian Ventura 
University of Córdoba
Córdoba, Spain

Min Zhou 
Noah's Ark Lab, Huawei
Beijing, China

Ilaria Bordino
UniCredit
Rome, Italy

Francesco Gullo 
Unicredit
Rome, Italy

Lorenzo Severini
Unicredit
Rome, Italy

Przemyslaw Biecek 
Warsaw University of Technology
Warsaw, Poland

Mykola Pechenizkiy 
Eindhoven University of Technology
Eindhoven, The Netherlands

Christopher Buckley 
University of Sussex
Brighton, UK

Pablo Lanillos 
Radboud University Nijmegen
Nijmegen, The Netherlands

Tim Verbelen 
Ghent University
Ghent, Belgium

Giuseppina Andresini 
University of Bari Aldo Moro
Bari, Italy

Ibéria Medeiros 
University of Lisbon
Lisbon, Portugal

M. Saqib Nawaz 
Harbin Institute of Technology
Harbin, China

Meng Sun 
Peking University
Beijing, China

Valerio Bitetta
UniCredit
Milan, Italy

Andrea Ferretti
UniCredit
Milan, Italy

Giovanni Ponti
ENEA Headquarters
Portici, Italy

Rita Ribeiro 
University of Porto
Porto, Portugal

João Gama 
University of Porto
Porto, Portugal

Lee Cooper 
Northwestern University
Chicago, IL, USA

Jonas Richiardi 
University of Lausanne
Lausanne, Switzerland

Diego Saldana Miranda 
F. Hoffmann–La Roche Ltd
Basel, Switzerland

Guilherme Graça 
University of Lisbon
Lisbon, Portugal

Ricard Gavalda 
UPC BarcelonaTech
Barcelona, Spain

Naghmeh Ghazaleh 
PD Personalised Healthcare
Basel, Switzerland

Damian Roqueiro 
ETH Zurich
Basel, Switzerland

Konstantinos Sechidis 
Novartis Pharma AG
Basel, Switzerland

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 2021, ECML PKDD was held virtually during September 13–17 due to the COVID-19 pandemic.

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. Workshop on Advances in Interpretable Machine Learning and Artificial Intelligence (AIMLAI 2021)
2. Workshop on Parallel, Distributed, and Federated Learning (PDFL 2021)
3. Workshop on Graph Embedding and Mining (GEM 2021)
4. Workshop on Machine Learning for Irregular Time Series (ML4ITS 2021)
5. Workshop on IoT, Edge, and Mobile for Embedded Machine Learning (ITEM 2021)
6. Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2021)
7. Workshop on Bias and Fairness in Artificial Intelligence (BIAS 2021)
8. Workshop on Workshop on Active Inference (IWAI 2021)
9. Workshop on Machine Learning for Cybersecurity (MLCS 2021)
10. Workshop on Machine Learning in Software Engineering (MLiSE 2021)
11. Workshop on Mining Data for Financial Applications (MIDAS 2021)
12. Workshop on Data Science for Social Good (SoGood 2021)
13. Workshop on Machine Learning for Pharma and Healthcare Applications (PharML 2021)
14. Workshop on Machine Learning for Buildings Energy Management (MLBEM 2021)

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 2021

Michael Kamp
on behalf of the volume editors

Organization

Workshop Chairs

AIMLAI 2021

Adrien Bibal	University of Namur, Belgium
Tassadit Bouadi	University of Rennes, France
Benoît Frénay	University of Namur, Belgium
Luis Galárraga	Inria, France
José Oramas	University of Antwerp, Belgium

PDFL 2021

Michael Kamp	IKIM, Ruhr-University Bochum, Germany/Monash University, Australia
Linara Adilova	Ruhr-University Bochum and Fraunhofer IAIS, Germany
Yamuna Krishnamurthy	Royal Holloway, University of London, UK

GEM 2021

Bo Kang	Ghent University, Belgium
Christine Largeron	Université Jean Monnet, France
Jefrey Lijffijt	Ghent University, Belgium
Tiphaine Viard	Telecom Paris, France
Pascal Welke	University of Bonn, Germany

ML4ITS 2021

Massimiliano Ruocco	Norwegian University of Science and Technology, Norway
Erlend Aune	BI Norwegian Business School, Norway
Claudio Gallicchio	University of Pisa, Italy

ITEM 2021

Gregor Schiele	University of Duisburg-Essen, Germany
Franz Pernkopf	Graz University of Technology, Austria
Michaela Blott	Xilinx Research, Ireland
Holger Fröning	Heidelberg University, Germany
Günther Schindler	Heidelberg University, Germany

XKDD 2021

Riccardo Guidotti	University of Pisa, Italy
Anna Monreale	University of Pisa, Italy
Salvatore Rinzivillo	ISTI-CNR, Italy
Przemyslaw Biecek	Warsaw University of Technology, Poland

BIAS 2021

Eirini Ntoutsi	Freie Universität Berlin, Germany
Mykola Pechenizkiy	Eindhoven University of Technology, The Netherlands
Bodo Rosenhahn	Leibniz Universität Hannover, Germany

IWAI 2021

Christopher Buckley	University of Sussex, UK
Daniela Cialfi	University of Chieti-Pescara, Italy
Pablo Lanillos	Donders Institute, The Netherlands
Maxwell Ramstead	McGill University, Canada
Tim Verbelen	Ghent University, Belgium

MLCS 2021

Pedro Ferreira	University of Lisbon, Portugal
Michael Kamp	IKIM, Ruhr-University Bochum, Germany/Monash University, Australia
Giuseppina Andresini	University of Bari Aldo Moro, Italy
Donato Malerba	University of Bari Aldo Moro, Italy
Ibéria Medeiros	University of Lisbon, Portugal

MLiSE 2021

Philippe Fournier-Viger	Shenzhen University, China
M. Saqib Nawaz	Harbin Institute of Technology, China
Sebastian Ventura	University of Cordoba, Spain
Meng Sun	Peking University, China
Min Zhou	Huawei, China

MIDAS 2021

Valerio Bitetta	Unicredit, Italy
Ilaria Bordino	Unicredit, Italy
Andrea Ferretti	Unicredit, Italy
Francesco Gullo	Unicredit, Italy
Giovanni Ponti	ENEA, Italy
Lorenzo Severini	Unicredit, Italy

SoGood 2021

Ricard Gavalda
Irena Koprinska
João Gama
Rita P. Ribeiro

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

PharML 2021

Lee Cooper
Naghmeh Ghazaleh
Jonas Richiardi

Damian Roqueiro
Diego Saldana
Konstantinos Sechidis

Northwestern University, USA
F. Hoffmann–La Roche Ltd., Switzerland
Lausanne University Hospital and University of
Lausanne, Switzerland
ETH Zurich and D-BSSE, Switzerland
F. Hoffmann–La Roche Ltd., Switzerland
Novartis Pharma AG, Switzerland

MLBEM 2021

Pedro Ferreira
Guilherme Graça

University of Lisbon, Portugal
University of Lisbon, Portugal

Contents – Part I

Advances in Interpretable Machine Learning and Artificial Intelligence

TS-MULE: Local Interpretable Model-Agnostic Explanations for Time Series Forecast Models	5
<i>Udo Schlegel, Duy Lam Vo, Daniel A. Keim, and Daniel Seebacher</i>	
Interpretable Models via Pairwise Permutations Algorithm	15
<i>Troy Maasland, João Pereira, Diogo Bastos, Marcus de Goffau, Max Nieuwdorp, Aeilko H. Zwinderman, and Evgeni Levin</i>	
A Classification of Anomaly Explanation Methods	26
<i>Véronne Yepmo Tchaghe, Grégory Smits, and Olivier Pivert</i>	
Bringing a Ruler Into the Black Box: Uncovering Feature Impact from Individual Conditional Expectation Plots	34
<i>Andrew Yeh and Anhthy Ngo</i>	
Reject and Cascade Classifier with Subgroup Discovery for Interpretable Metagenomic Signatures	49
<i>Maxence Queyrel, Alexandre Templier, and Jean-Daniel Zucker</i>	
Demystifying Graph Neural Network Explanations	67
<i>Anna Himmelhuber, Mitchell Joblin, Martin Ringsquandl, and Thomas Runkler</i>	
On the Transferability of Neural Models of Morphological Analogies	76
<i>Safa Alsaidi, Amandine Decker, Puthineath Lay, Esteban Marquer, Pierre-Alexandre Murena, and Miguel Couceiro</i>	
Behavior of k-NN as an Instance-Based Explanation Method	90
<i>Chhavi Yadav and Kamalika Chaudhuri</i>	
Enhancing Performance of Occlusion-Based Explanation Methods by a Hierarchical Search Method on Input Images	97
<i>Hamed Behzadi-Khormouji and Habib Rostami</i>	
Post-hoc Counterfactual Generation with Supervised Autoencoder	105
<i>Victor Guyomard, Françoise Fessant, Tassadit Bouadi, and Thomas Guyet</i>	

Parallel, Distributed, and Federated Learning

Differentially Private Learning from Label Proportions	119
<i>Timon Sachweh, Daniel Boiar, and Thomas Liebig</i>	
Approaches to Uncertainty Quantification in Federated Deep Learning	128
<i>Florian Linsner, Linara Adilova, Sina Däubener, Michael Kamp, and Asja Fischer</i>	
Optimized Federated Learning on Class-Biased Distributed Data Sources	146
<i>Yongli Mou, Jiahui Geng, Sascha Welten, Chunming Rong, Stefan Decker, and Oya Beyan</i>	
Splitting Algorithms for Federated Learning	159
<i>Saber Malekmohammadi, Kiarash Shaloudegi, Zeou Hu, and Yaoliang Yu</i>	
Migrating Models: A Decentralized View on Federated Learning	177
<i>Péter Kiss and Tomáš Horváth</i>	

Graph Embedding and Mining

The Effects of Randomness on the Stability of Node Embeddings	197
<i>Tobias Schumacher, Hinrikus Wolf, Martin Ritzert, Florian Lemmerich, Martin Grohe, and Markus Strohmaier</i>	
Graph Homomorphism Features: Why Not Sample?	216
<i>Paul Beaujean, Florian Sikora, and Florian Yger</i>	
Neural Maximum Independent Set	223
<i>Thomas Pontoizeau, Florian Sikora, Florian Yger, and Tristan Cazenave</i>	
Fea2Fea: Exploring Structural Feature Correlations via Graph Neural Networks	238
<i>Jiaqing Xie and Rex Ying</i>	
Web Image Context Extraction with Graph Neural Networks and Sentence Embeddings on the DOM Tree	258
<i>Chen Dang, Hicham Randrianarivo, Raphaël Fournier-S'niehotta, and Nicolas Audebert</i>	
Towards Mining Generalized Patterns from RDF Data and a Domain Ontology	268
<i>Tomas Martin, Victor Fuentes, Petko Valtchev, Abdoulaye Baniré Diallo, René Lacroix, Maxime Leduc, and Mounir Boukadoum</i>	

Machine Learning for Irregular Time Series

Homological Time Series Analysis of Sensor Signals from Power Plants	283
<i>Luciano Melodia and Richard Lenz</i>	

Continuous-Discrete Recurrent Kalman Networks for Irregular Time Series	300
<i>Mona Schirmer, Mazin Eltayeb, and Maja Rudolph</i>	

Adversarial Generation of Temporal Data: A Critique on Fidelity of Synthetic Data	306
<i>Ankur Debnath, Nitish Gupta, Govind Waghmare, Hardik Wadhwa, Siddhartha Asthana, and Ankur Arora</i>	

IoT, Edge, and Mobile for Embedded Machine Learning

Towards Precomputed 1D-Convolutional Layers for Embedded FPGAs	327
<i>Lukas Einhaus, Chao Qian, Christopher Ringhofer, and Gregor Schiele</i>	

Embedded Face Recognition for Personalized Services in the Assistive Robotics	339
<i>Iris Walter, Jonas Ney, Tim Hotfilter, Vladimir Rybalkin, Julian Hoefer, Norbert Wehn, and Jürgen Becker</i>	

FLight: FPGA Acceleration of <u>Lightweight</u> DNN Model Inference in Industrial Analytics	351
<i>Hassan Ghasemzadeh Mohammadi, Felix Paul Jentzsch, Maurice Kuschel, Rahil Arshad, Sneha Rautmare, Suraj Manjunatha, Marco Platzner, Alexander Boschmann, and Dirk Schollbach</i>	

Exploring Cell-Based Neural Architectures for Embedded Systems	363
<i>Ilja van Ipenburg, Dolly Sapra, and Andy D. Pimentel</i>	

Design Space Exploration of Time, Energy, and Error Rate Trade-offs for CNNs Using Accuracy-Programmable Instruction Set Processors	375
<i>Armin Schuster, Christian Heidorn, Marcel Brand, Oliver Keszocze, and Jürgen Teich</i>	

Ultra-low Power Machinery Fault Detection Using Deep Neural Networks	390
<i>Sven Nitzsche, Moritz Neher, Stefan von Dosky, and Jürgen Becker</i>	

SPNC: Fast Sum-Product Network Inference	397
<i>Lukas Sommer, Cristian Axenie, and Andreas Koch</i>	

Towards Addressing Noise and Static Variations of Analog Computations Using Efficient Retraining	409
<i>Bernhard Klein, Lisa Kuhn, Johannes Weis, Arne Emmel, Yannik Stradmann, Johannes Schemmel, and Holger Fröning</i>	

eXplainable Knowledge Discovery in Data Mining

The Next Frontier: AI We Can Really Trust	427
<i>Andreas Holzinger</i>	
This Looks Like That, Because ... Explaining Prototypes for Interpretable Image Recognition	441
<i>Meike Nauta, Annemarie Jutte, Jesper Provoost, and Christin Seifert</i>	
Prototypical Convolutional Neural Network for a Phrase-Based Explanation of Sentiment Classification	457
<i>Kamil Pluciński, Mateusz Lango, and Jerzy Stefanowski</i>	
Explanations for Network Embedding-Based Link Predictions	473
<i>Bo Kang, Jefrey Lijffijt, and Tijl De Bie</i>	
Exploring Counterfactual Explanations for Classification and Regression Trees	489
<i>Suryabhan Singh Hada and Miguel Á. Carreira-Perpiñán</i>	
Towards Explainable Meta-learning	505
<i>Katarzyna Woźnica and Przemysław Biecek</i>	
How to Choose an Explainability Method? Towards a Methodical Implementation of XAI in Practice	521
<i>Tom Vermeire, Thibault Laugel, Xavier Renard, David Martens, and Marcin Detryniecki</i>	
Using Explainable Boosting Machines (EBMs) to Detect Common Flaws in Data	534
<i>Zhi Chen, Sarah Tan, Harsha Nori, Kori Inkpen, Yin Lou, and Rich Caruana</i>	

Bias and Fairness in AI

Algorithmic Factors Influencing Bias in Machine Learning	559
<i>William Blanzeisky and Pádraig Cunningham</i>	

Desiderata for Explainable AI in Statistical Production Systems of the European Central Bank	575
<i>Carlos Mougán Navarro, Georgios Kanellos, and Thomas Gottron</i>	
Robustness of Fairness: An Experimental Analysis	591
<i>Serafina Kamp, Andong Luis Li Zhao, and Sindhu Kuttty</i>	
Co-clustering for Fair Recommendation	607
<i>Gabriel Frisch, Jean-Benoist Leger, and Yves Grandvalet</i>	
Learning a Fair Distance Function for Situation Testing	631
<i>Daphne Lenders and Toon Calders</i>	
Towards Fairness Through Time	647
<i>Alessandro Castelnovo, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Andrea Cosentini</i>	
International Workshop on Active Inference	
Active Inference for Stochastic Control	669
<i>Aswin Paul, Noor Sajid, Manoj Gopalkrishnan, and Adeel Razi</i>	
Towards Stochastic Fault-Tolerant Control Using Precision Learning and Active Inference	681
<i>Mohamed Baioumy, Corrado Pezzato, Carlos Hernández Corbato, Nick Hawes, and Riccardo Ferrari</i>	
On the Convergence of DEM's Linear Parameter Estimator	692
<i>Ajith Anil Meera and Martijn Wisse</i>	
Disentangling What and Where for 3D Object-Centric Representations Through Active Inference	701
<i>Toon Van de Maele, Tim Verbelen, Ozan Çatal, and Bart Dhoedt</i>	
Rule Learning Through Active Inductive Inference	715
<i>Tore Erdmann and Christoph Mathys</i>	
Interpreting Dynamical Systems as Bayesian Reasoners	726
<i>Nathaniel Virgo, Martin Biehl, and Simon McGregor</i>	
Blankets All the Way up – the Economics of Active Inference	763
<i>Morten Henriksen</i>	
Filtered States: Active Inference, Social Media and Mental Health	772
<i>Ben White and Mark Miller</i>	

Ideas Worth Spreading: A Free Energy Proposal for Cumulative Cultural Dynamics	784
<i>Natalie Kastel and Casper Hesp</i>	
Dream to Explore: 5-HT2a as Adaptive Temperature Parameter for Sophisticated Affective Inference	799
<i>Adam Safron and Zahra Sheikhbahae</i>	
Inferring in Circles: Active Inference in Continuous State Space Using Hierarchical Gaussian Filtering of Sufficient Statistics	810
<i>Peter Thestrup Waade, Nace Mikus, and Christoph Mathys</i>	
On Solving a Stochastic Shortest-Path Markov Decision Process as Probabilistic Inference	819
<i>Mohamed Baioumy, Bruno Lacerda, Paul Duckworth, and Nick Hawes</i>	
Habitual and Reflective Control in Hierarchical Predictive Coding	830
<i>Paul F. Kinghorn, Beren Millidge, and Christopher L. Buckley</i>	
Deep Active Inference for Pixel-Based Discrete Control: Evaluation on the Car Racing Problem	843
<i>N. T. A. van Hoeffelen and Pablo Lanillos</i>	
Robot Localization and Navigation Through Predictive Processing Using LiDAR	857
<i>Daniel Burghardt and Pablo Lanillos</i>	
Sensorimotor Visual Perception on Embodied System Using Free Energy Principle	865
<i>Kanako Esaki, Tadayuki Matsumura, Kiyoto Ito, and Hiroyuki Mizuno</i>	
Author Index	879

Contents – Part II

Machine Learning for CyberSecurity

Dealing with Imbalanced Data in Multi-class Network Intrusion Detection Systems Using XGBoost	5
<i>Malik AL-Essa and Annalisa Appice</i>	
Adversarial Robustness of Probabilistic Network Embedding for Link Prediction	22
<i>Xi Chen, Bo Kang, Jefrey Liijffijt, and Tijl De Bie</i>	
Practical Black Box Model Inversion Attacks Against Neural Nets	39
<i>Thomas Bekman, Masoumeh Abolfathi, Haadi Jafarian, Ashis Biswas, Farnoush Banaei-Kashani, and Kuntal Das</i>	
NBcoded: Network Attack Classifiers Based on Encoder and Naive Bayes Model for Resource Limited Devices	55
<i>Lander Seguro-Gil, Francesco Zola, Xabier Echeberria-Barrio, and Raul Orduna-Urrutia</i>	

Workshop on Machine Learning in Software Engineering

A Stacked Bidirectional LSTM Model for Classifying Source Codes Built in MPLs	75
<i>Md. Mostafizer Rahman, Yutaka Watanobe, Rage Uday Kiran, and Raihan Kabir</i>	
Metamorphic Malware Behavior Analysis Using Sequential Pattern Mining	90
<i>M. Saqib Nawaz, Philippe Fournier-Viger, M. Zohaib Nawaz, Guoting Chen, and Youxi Wu</i>	
Applying Machine Learning to Risk Assessment in Software Projects	104
<i>André Sousa, João Pascoal Faria, João Mendes-Moreira, Duarte Gomes, Pedro Castro Henriques, and Ricardo Graça</i>	
SampleFix: Learning to Generate Functionally Diverse Fixes	119
<i>Hossein Hajipour, Apratim Bhattacharyya, Cristian-Alexandru Staicu, and Mario Fritz</i>	
Linguistic Analysis of Stack Overflow Data: Native English vs Non-native English Speakers	134
<i>Janneke Morin and Krishnendu Ghosh</i>	

IReEn: Reverse-Engineering of Black-Box Functions via Iterative Neural Program Synthesis	143
<i>Hossein Hajipour, Mateusz Malinowski, and Mario Fritz</i>	
Machine Learning for Intelligent Industrial Design	158
<i>Philippe Fournier-Viger, M. Saqib Nawaz, Wei Song, and Wensheng Gan</i>	
MIning DAta for financial applicationS	
Financial Forecasting with Word Embeddings Extracted from News: A Preliminary Analysis	179
<i>Luca Barbaglia, Sergio Consoli, and Susan Wang</i>	
On Neural Forecasting and News Emotions: The Case of the Spanish Stock Market	189
<i>Sergio Consoli, Matteo Negri, Amirhossein Tebbifakhr, Elisa Tosetti, and Marco Turchi</i>	
Adaptive Supervised Learning for Financial Markets Volatility Targeting Models	195
<i>Eric Benhamou, David Saliel, Serge Tabachnik, Corentin Bourdeix, François Chareyron, and Beatrice Guez</i>	
Efficient Analysis of Transactional Data Using Graph Convolutional Networks	210
<i>Hamish Hall, Pedro Baiz, and Philip Nadler</i>	
A Reasoning Approach to Financial Data Exchange with Statistical Confidentiality	226
<i>Luigi Bellomarini, Livia Blasi, Rosario Laurendi, and Emanuel Sallinger</i>	
Forecasting Longevity for Financial Applications: A First Experiment with Deep Learning Methods	232
<i>Jorge M. Bravo</i>	
Sixth Workshop on Data Science for Social Good (SoGood 2021)	
Ensuring the Inclusive Use of NLP in the Global Response to COVID-19	259
<i>Alexandra Sasha Luccioni, Katherine Hoffmann Pham, Cynthia Sin Nga Lam, Joseph Aylett-Bullock, and Miguel Luengo-Oroz</i>	
A Framework for Building <i>pro-bono</i> Data for Good Projects	267
<i>Miguel José Monteiro and Paulo Maia</i>	

Improving Smart Waste Collection Using AutoML	283
<i>Sónia Teixeira, Guilherme Londres, Bruno Veloso, Rita P. Ribeiro, and João Gama</i>	
Applying Machine Learning for Traffic Forecasting in Porto, Portugal	299
<i>Paulo Maia, Joana Morgado, Tiago Gonçalves, and Tomé Albuquerque</i>	
IRLCov19: A Large COVID-19 Multilingual Twitter Dataset of Indian Regional Languages	309
<i>Deepak Uniyal and Amit Agarwal</i>	
Automated ESG Report Analysis by Joint Entity and Relation Extraction	325
<i>Adrien Ehrhardt and Minh Tuan Nguyen</i>	
Hate Speech Detection in Clubhouse	341
<i>Hadi Mansourifar, Dana Alsagheer, Reza Fathi, Weidong Shi, Lan Ni, and Yan Huang</i>	
Error Variance, Fairness, and the Curse on Minorities	352
<i>Emma Beauxis-Aussalet</i>	
Machine Learning for Pharma and Healthcare Applications	
Embed Wisely: An Ensemble Approach to Predict ICD Coding	371
<i>Pavithra Rajendran, Alexandros Zenonos, Joshua Spear, and Rebecca Pope</i>	
L₁-Regularized Neural Ranking for Risk Stratification and Its Application to Prediction of Time to Distant Metastasis in Luminal Node Negative Chemotherapy Naïve Breast Cancer Patients	390
<i>Fayyaz Minhas, Michael S. Toss, Noor ul Wahab, Emad Rakha, and Nasir M. Rajpoot</i>	
Extracting Multilingual Relations with Joint Learning of Language Models	401
<i>Nuria García-Santa and Kendrick Cetina</i>	
Supervised Clustering for Subgroup Discovery: An Application to COVID-19 Symptomatology	408
<i>Aidan Cooper, Orla Doyle, and Alison Bourke</i>	
Diagnostic Surveillance of High-Grade Gliomas: Towards Automated Change Detection Using Radiology Report Classification	423
<i>Tommaso Di Noto, Chirine Atat, Eduardo Gamito Teiga, Monika Hegi, Andreas Hottinger, Meritxell Bach Cuadra, Patric Hagmann, and Jonas Richiardi</i>	

All You Need is Color: Image Based Spatial Gene Expression Prediction Using Neural Stain Learning	437
<i>Muhammad Dawood, Kim Branson, Nasir M. Rajpoot, and Fayyaz ul Amir Afsar Minhas</i>	
Explaining a Random Survival Forest by Extracting Prototype Rules	451
<i>Klest Dedja, Felipe Kenji Nakano, Konstantinos Pliakos, and Celine Vens</i>	
Federated Learning of Oligonucleotide Drug Molecule Thermodynamics with Differentially Private ADMM-Based SVM	459
<i>Shirin Tavara, Alexander Schliep, and Debabrota Basu</i>	
Mining Medication-Effect Relations from Twitter Data Using Pre-trained Transformer Language Model	468
<i>Keyuan Jiang, Dingkai Zhang, and Gordon R. Bernard</i>	
A Scalable AI Approach for Clinical Trial Cohort Optimization	479
<i>Xiong Liu, Cheng Shi, Uday Deore, Yingbo Wang, Myah Tran, Iya Khalil, and Murthy Devarakonda</i>	
Private Cross-Silo Federated Learning for Extracting Vaccine Adverse Event Mentions	490
<i>Pallika Kanani, Virendra J. Marathe, Daniel Peterson, Rave Harpaz, and Steve Bright</i>	
Machine Learning for Buildings Energy Management	
Building Appliances Energy Performance Assessment	511
<i>Zygmantas Jasiunas, Pedro M. Ferreira, and José Cecílio</i>	
Drilling a Large Corpus of Document Images of Geological Information Extraction	525
<i>Jean-Louis Debezia, Mélodie Boillet, Christopher Kermorvant, and Quentin Barral</i>	
Data Envelopment Analysis for Energy Audits of Housing Properties	531
<i>Sushodhan Vaishampayan, Aditi Pawde, Akshada Shinde, Manoj Apte, and Girish Keshav Palshikar</i>	
A Comparative Study on Machine Learning Algorithms for Assessing Energy Efficiency of Buildings	546
<i>Christian Nnaemeka Egwim, Oluwapelumi Oluwaseun Egunjobi, Alvaro Gomes, and Hafiz Alaka</i>	

Buildings Occupancy Estimation: Preliminary Results Using Bluetooth
Signals and Artificial Neural Networks 567
 Frederico Apolónia, Pedro M. Ferreira, and José Cecílio

Author Index 581