

Robert Wrembel  
Johann Gamper  
Gabriele Kotsis  
A Min Tjoa  
Ismail Khalil (Eds.)

LNCS 14148

# Big Data Analytics and Knowledge Discovery

25th International Conference, DaWaK 2023  
Penang, Malaysia, August 28–30, 2023  
Proceedings

**DaWak 2023**



Springer

# Lecture Notes in Computer Science

14148

## Founding Editors


Gerhard Goos  
Juris Hartmanis

## Editorial Board Members

Elisa Bertino, *Purdue University, West Lafayette, IN, USA*

Wen Gao, *Peking University, Beijing, China*

Bernhard Steffen , *TU Dortmund University, Dortmund, Germany*

Moti Yung , *Columbia University, New York, NY, USA*

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.


LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Robert Wrembel · Johann Gamper ·  
Gabriele Kotsis · A Min Tjoa · Ismail Khalil  
Editors

# Big Data Analytics and Knowledge Discovery

25th International Conference, DaWaK 2023  
Penang, Malaysia, August 28–30, 2023  
Proceedings

*Editors*

Robert Wrembel   
Poznań University of Technology  
Poznan, Poland

Johann Gamper   
Free University of Bozen-Bolzano  
Bozen-Bolzano, Italy

Gabriele Kotsis  
Johannes Kepler University Linz  
Linz, Austria

A Min Tjoa   
Vienna University of Technology  
Vienna, Austria

Ismail Khalil  
Johannes Kepler University Linz  
Linz, Austria

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Computer Science

ISBN 978-3-031-39830-8

ISBN 978-3-031-39831-5 (eBook)

<https://doi.org/10.1007/978-3-031-39831-5>

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

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

# Preface

It is with great pleasure that we introduce the proceedings of the 25th International Conference on Big Data Analytics and Knowledge Discovery (DAWAK 2023). This conference brought together leading experts, researchers, and practitioners from around the world to exchange ideas, share insights, and push the boundaries of knowledge in the fields of big data, data analytics, artificial intelligence, and machine learning. The compilation of papers presented in this volume represents a diverse range of topics and showcases the cutting-edge research that was discussed during the conference.

In today's data-driven world, the proliferation of vast and complex datasets has necessitated the development of advanced techniques and methods to extract meaningful insights and to discover valuable knowledge. The advent of big data analytics has transformed the way we process, analyze, and interpret data, empowering organizations and individuals to make informed decisions and gain a competitive edge. Coupled with the power of artificial intelligence and machine learning, these technologies have opened up new frontiers in data exploration, predictive modeling, and automated decision-making.

DAWAK 2023 served as a hub for interdisciplinary collaboration, fostering the exchange of ideas between researchers, practitioners, and industry experts. The papers included in this volume represent the culmination of countless hours of dedicated work, pushing the boundaries of what is possible in the realms of data analytics, artificial intelligence, and machine learning.

The breadth of topics covered in this collection is truly remarkable, encompassing areas such as data mining, pattern recognition, natural language processing, recommendation systems, deep learning, and reinforcement learning, among others. Each paper has undergone an average of 3 single-blind reviews in a rigorous peer-review process, ensuring the highest quality and relevance to the conference's theme.

We are proud to report authors from more than 22 different countries submitted papers to DAWAK this year. Our program committees have conducted close to 300 reviews. From 83 submitted papers the program committee decided to accept 18 full papers with an acceptance rate of 20%, a rate lower than previous DAWAK conferences.

This year again the top DAWAK 2023 papers will be invited to a special issue of the *Data & Knowledge Engineering* journal. Here we express our gratitude to Carson Woo, the Editor in-Chief of the journal, for hosting these DAWAK 2023 papers in the special issue.

We would like to extend our gratitude to all the authors who have contributed their valuable research to this volume. Their passion, expertise, and dedication to their respective fields have enriched the scientific community and paved the way for groundbreaking discoveries.

We also want to express our appreciation to the program committee members, whose expertise and thorough evaluation played an instrumental role in selecting and refining the papers for inclusion.

Furthermore, we would like to extend our thanks to the conference organizers, keynote speakers, and attendees for their unwavering support and enthusiasm. Their commitment to advancing knowledge and fostering collaboration has been instrumental in creating an environment where innovation flourishes and ideas thrive.

As you peruse the pages of these proceedings, we hope that you find inspiration in the remarkable research and advancements showcased within. Whether you are a seasoned researcher, a student, or an industry professional, we believe that the insights shared here will not only broaden your understanding but also inspire you to embark on your own journey of discovery.

Finally, we would like to express our heartfelt appreciation to the entire conference community. It is your collective efforts and unwavering dedication that have made this event a resounding success, enabling us to push the boundaries of big data analytics, knowledge discovery, artificial intelligence, and machine learning.

August 2023

Robert Wrembel  
Johann Gamper  
Gabriele Kotsis  
A Min Tjoa  
Ismail Khalil

# Organization

## Program Committee Chairs

Robert Wrembel  
Johann Gamper

Poznań University of Technology, Poland  
Free University of Bozen-Bolzano, Italy

## Steering Committee

Gabriele Kotsis  
A Min Tjoa  
Robert Wille  
Bernhard Moser  
Ismail Khalil

Johannes Kepler University Linz, Austria  
Vienna University of Technology, Austria  
Software Competence Center Hagenberg, Austria  
Software Competence Center Hagenberg, Austria  
Johannes Kepler University Linz, Austria

## Program Committee Members

Alberto Abello  
Cristina Aguiar  
Lars Ailo Bongo  
Syed Muhammad Fawad Ali  
Witold Andrzejewski  
Faten Atigui  
Sylvio Barbon Junior  
Ladjel Bellatreche  
Soumia Benkrid

Universitat Politècnica de Catalunya, Spain  
Universidade de São Paulo, Brazil  
University of Tromsø, Norway  
Accenture, Germany  
Poznań University of Technology, Poland  
CNAM, France  
University of Trieste, Italy  
LIAS/ENSMA, France  
Ecole Nationale Supérieure d'Informatique,  
Algeria

Fadila Bentayeb  
Jorge Bernardino  
Vasudha Bhatnagar  
Besim Bilalli  
Sandro Bimonte  
Pawel Boinski  
Kamel Boukhalfa  
Omar Boussaid  
Stephane Bressan  
Tiziana Catarci

Université Lumière Lyon 2, France  
Polytechnic Institute of Coimbra, Portugal  
University of Delhi, India  
Universitat Politècnica de Catalunya, Spain  
INRAE, France  
Poznań University of Technology, Poland  
USTHB, Algeria  
Université Lumière Lyon 2, France  
National University of Singapore, Singapore  
Sapienza University of Rome, Italy



Sharma Chakravarthy	University of Texas at Arlington, USA
Silvia Chiusano	Politecnico di Torino, Italy
Frans Coenen	University of Liverpool, UK
Philippe Cudré-Mauroux	Universität Freiburg, Switzerland
Laurent d’Orazio	Université de Rennes 1, CNRS IRISA, France
Jérôme Darmont	Université Lumière Lyon 2, France
Soumyava Das	Teradata Lab, USA
Karen Davis	Miami University, USA
Anton Dignös	Free University of Bozen-Bolzano, Italy
Christos Doulkeridis	University of Piraeus, Greece
Markus Endres	University of Passau, Germany
Leonidas Fegaras	University of Texas at Arlington, USA
Matteo Francia	University of Bologna, Italy
Filippo Furfaro	University of Calabria, Italy
Pedro Furtado	University of Coimbra, Portugal
Luca Gagliardelli	University of Modena and Reggio Emilia, Italy
Enrico Gallinucci	University of Bologna, Italy
Kazuo Goda	University of Tokyo, Japan
Matteo Golfarelli	University of Bologna, Italy
Marcin Gorawski	Silesian University of Technology, Poland
Sven Groppe	University of Lübeck, Germany
Frank Höppner	Ostfalia University of Applied Sciences, Germany
Mirjana Ivanovic	University of Novi Sad, Serbia
Stéphane Jean	LIAS/ISAE-ENSMA and University of Poitiers, France
Selma Khouri	Ecole Nationale Supérieure d’Informatique, Algeria
Kyoung-Sook Kim	National Institute of Advanced Industrial Science and Technology (AIST), Japan
Young-Kuk Kim	Chungnam National University, South Korea
Nicolas Labroche	University of Tours, France
Jens Lechtenbörger	University of Münster, Germany
Young-Koo Lee	Kyung Hee University, South Korea
Carson Leung	University of Manitoba, Canada
Sebastian Link	University of Auckland, New Zealand
Woong-Kee Loh	Gachon University, South Korea
Sofian Maabout	LaBRI/University of Bordeaux, France
Patrick Marcel	Université de Tours, France
Alex Mircoli	Polytechnic University of Marche, Italy
Jun Miyazaki	Tokyo Institute of Technology, Japan
Rim Moussa	ENICarthage, Tunisia
Sergi Nadal	Universitat Politècnica de Catalunya, Spain

Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Boris Novikov	National Research University Higher School of Economics, Russia
Hiroaki Ohshima	University of Hyogo, Japan
Makoto Onizuka	Osaka University, Japan
Carlos Ordóñez	University of Houston, USA
Veronika Peralta	University of Tours, France
Franck Ravat	IRIT/Université Toulouse 1, France
Ilya Safro	University of Delaware, USA
Mahmoud Sakr	Université Libre de Bruxelles, Belgium
Abhishek Santra	University of Texas at Arlington, USA
Sana Sellami	Aix-Marseille University, France
Alkis Simitsis	HP Labs, USA
Darja Solodovnikova	University of Latvia, Latvia
Nadine Steinmetz	TU Ilmenau, Germany
Emanuele Storti	Università Politecnica delle Marche, Italy
Olivier Teste	IRIT, France
Dimitri Theodoratos	New Jersey Institute of Technology, USA
Maik Thiele	Technische Universität Dresden, Germany
Chrisa Tsinaraki	Technical University of Crete, Greece
Alejandro Vaisman	Instituto Tecnológico de Buenos Aires, Argentina
Luca Zecchini	University of Modena and Reggio Emilia, Italy
Yongjun Zhu	Yonsei University, South Korea

## External Reviewers

Simone Agostinelli	Sapienza University of Rome, Italy
Guessoum Meriem Amel	University of Science and Technology Houari Boumediene, Algeria
Bharti Bharti	University of Delhi, India
Tobias Bleifuss	Hasso-Plattner Institute, Germany
Abdesslam Chalallah	UMR Herbivores, INRAE, France
Renukswamy Chikkamath	Munich University of Applied Sciences, Germany
Gianluca Cima	Sapienza University of Rome, Italy
Daniele Cono D'Elia	Sapienza University of Rome, Italy
Deghmani Faiza	University of Science and Technology Houari Boumediene, Algeria
Chiara Forresi	University of Bologna, Italy
Nguyen Hoang Hai	University of Manitoba, Canada
Vikas Kumar	University of Delhi, India

Maryam Mozaffari  
Toshpulatov Mukhiddin  
Christian Napoli  
Anifat Olawayin  
Adam Pazdor  
Kassem Sabeh  
Ionel Eduard Stan  
Tung Son Tran

Free University of Bozen-Bolzano, Italy  
Inha University, South Korea  
Sapienza University of Rome, Italy  
University of Manitoba, Canada  
University of Manitoba, Canada  
Free University of Bozen-Bolzano, Italy  
Free University of Bozen-Bolzano, Italy  
Munich University of Applied Sciences, Germany

## Organizers



# From an Interpretable Predictive Model to a Model Agnostic Explanation (Abstract of Keynote Talk)

Osmar R. Zaiane

Amii Fellow and Canada CIFAR AI Chair, University of Alberta, Canada

**Abstract.** Today, the limelight is on Deep Learning. With the massive success of deep learning, other machine learning paradigms have had to take a backseat. Yet other models, particularly rule-based learning methods, are more readable and explainable and can even be competitive when labelled data is not abundant, and therefore could be more suitable for some applications where transparency is a must. One such rule-based method is the less-known Associative Classifier. The power of associative classifiers is to determine patterns from the data and perform classification based on the features most indicative of prediction. Early approaches suffer from cumbersome thresholds requiring prior knowledge. We present a new associative classifier approach that is even more accurate while generating a smaller model. It can also be used in an explainable AI pipeline to explain inferences from other classifiers, irrespective of the predictive model used inside the black box.

# Contents

## Data Quality

Using Ontologies as Context for Data Warehouse Quality Assessment .....	3
<i>Camila Sanz and Adriana Marotta</i>	
Preventing Technical Errors in Data Lake Analyses with Type Theory .....	18
<i>Alexis Guyot, Éric Leclercq, Annabelle Gillet, and Nadine Cullot</i>	
EXOS: Explaining Outliers in Data Streams .....	25
<i>Egawati Panjei and Le Gruenwald</i>	
Motif Alignment for Time Series Data Augmentation .....	42
<i>Omar Bahri, Peiyu Li, Soukaina Filali Boubrahimi, and Shah Muhammad Hamdi</i>	
State-Transition-Aware Anomaly Detection Under Concept Drifts .....	49
<i>Bin Li and Emmanuel Müller</i>	
Anomaly Detection in Financial Transactions Via Graph-Based Feature Aggregations .....	64
<i>Hewen Wang, Renchi Yang, and Jieming Shi</i>	
The Synergies of Context and Data Aging in Recommendations .....	80
<i>Anna Dalla Vecchia, Niccolò Marastoni, Barbara Oliboni, and Elisa Quintarelli</i>	

## Advanced Analytics and Pattern Discovery

Hypergraph Embedding Based on Random Walk with Adjusted Transition Probabilities .....	91
<i>Kazuya Nagasato, Satoshi Takabe, and Kazuyuki Shudo</i>	
Contextual Shift Method (CSM) .....	101
<i>Gernot Schmitz, Daniel Wilmes, Alexander Gerharz, Daniel Horn, and Emmanuel Müller</i>	
Utility-Oriented Gradual Itemsets Mining Using High Utility Itemsets Mining .....	107
<i>Audrey Fongue, Jerry Lonlac, and Norbert Tsopze</i>	

Discovery of Contrast Itemset with Statistical Background Between Two Continuous Variables .....	114
<i>Ke Zhang, Shogo Matsuno, and Shota Saito</i>	
DBGAN: A Data Balancing Generative Adversarial Network for Mobility Pattern Recognition .....	120
<i>Ke Zhang, Hengchang Liu, and Siobhán Clarke</i>	
Bitwise Vertical Mining of Minimal Rare Patterns .....	135
<i>Elieser Capillar, Chowdhury Abdul Mumin Ishmam, Carson K. Leung, Adam G. M. Pazdor, Prabhanshu Shrivastava, and Ngoc Bao Chau Truong</i>	
Inter-item Time Intervals in Sequential Patterns .....	142
<i>Thomas Kastner, Hubert Cardot, and Dominique H. Li</i>	
Fair-DSP: Fair Dynamic Survival Prediction on Longitudinal Electronic Health Record .....	149
<i>Xin Huang, Xiangyang Meng, Ni Zhao, Wenbin Zhang, and Jianwu Wang</i>	
<b>Machine Learning</b>	
DAT@Z21: A Comprehensive Multimodal Dataset for Rumor Classification in Microblogs .....	161
<i>Abderrazek Azri, Cécile Favre, Nouria Harbi, Jérôme Darmont, and Camille Noûs</i>	
Dealing with Data Bias in Classification: Can Generated Data Ensure Representation and Fairness? .....	176
<i>Manh Khoi Duong and Stefan Conrad</i>	
Random Hypergraph Model Preserving Two-Mode Clustering Coefficient .....	191
<i>Rikuya Miyashita, Kazuki Nakajima, Mei Fukuda, and Kazuyuki Shudo</i>	
A Non-overlapping Community Detection Approach Based on $\alpha$ -Structural Similarity .....	197
<i>Motaz Ben Hassine, Saïd Jabbour, Mourad Kmimech, Badran Raddaoui, and Mohamed Graïet</i>	
Improving Stochastic Gradient Descent Initializing with Data Summarization .....	212
<i>Robin Varghese and Carlos Ordonez</i>	

Feature Analysis of Regional Behavioral Facilitation Information Based on Source Location and Target People in Disaster .....	224
<i>Kosuke Wakasugi, Futo Yamamoto, Yu Suzuki, and Akiyo Nadamoto</i>	

Exploring Dialog Act Recognition in Open Domain Conversational Agents ....	233
<i>Maliha Sultana and Osmar R. Zaiane</i>	

UniCausal: Unified Benchmark and Repository for Causal Text Mining .....	248
<i>Fiona Anting Tan, Xinyu Zuo, and See-Kiong Ng</i>	

## Deep Learning

Accounting for Imputation Uncertainty During Neural Network Training .....	265
<i>Thomas Ranvier, Haytham Elghazel, Emmanuel Coquery, and Khalid Benabdeslem</i>	

Supervised Hybrid Model for Rumor Classification: A Comparative Study of Machine and Deep Learning Approaches .....	281
<i>Mehzabin Sadat Aothoi, Samin Ahsan, Najeefa Nikhat Choudhury, and Annajiat Alim Rasel</i>	

Attention-Based Counterfactual Explanation for Multivariate Time Series .....	287
<i>Peiyu Li, Omar Bahri, Soukaïna Filali Boubrahimi, and Shah Muhammad Hamdi</i>	

DRUM: A Real Time Detector for Regime Shifts in Data Streams via an Unsupervised, Multivariate Framework .....	294
<i>Adnan Bashir and Trilce Estrada</i>	

Hierarchical Graph Neural Network with Cross-Attention for Cross-Device User Matching .....	303
<i>Ali Taghibakhshi, Mingyuan Ma, Ashwath Aithal, Onur Yilmaz, Haggai Maron, and Matthew West</i>	

## Data Management

Unified Views for Querying Heterogeneous Multi-model Polystores .....	319
<i>Lea El Ahdab, Olivier Teste, Imen Megdiche, and Andre Peninou</i>	

RODD: Robust Outlier Detection in Data Cubes .....	325
<i>Lara Kuhlmann, Daniel Wilmes, Emmanuel Müller, Markus Pauly, and Daniel Horn</i>	

Data-driven and On-Demand Conceptual Modeling ..... 340  
*Damianos Chatziantoniou and Verena Kantere*

FLOWER: Viewing Data Flow in ER Diagrams ..... 356  
*Elijah Mitchell, Nabila Berkani, Ladjel Bellatreche, and Carlos Ordonez*

Supporting Big Healthcare Data Management and Analytics: The  
Cloud-Based QFLS Framework ..... 372  
*Alfredo Cuzzocrea and Selim Soufargi*

Beyond Traditional Flare Forecasting: A Data-driven Labeling Approach  
for High-fidelity Predictions ..... 380  
*Jinsu Hong, Anli Ji, Chetraj Pandey, and Berkay Aydin*

HKS: Efficient Data Partitioning for Stateful Streaming ..... 386  
*Adeel Aslam, Giovanni Simonini, Luca Gagliardelli, Angelo Mozzillo,  
and Sonia Bergamaschi*

A Fine-Grained Structural Partitioning Approach to Graph Compression ..... 392  
*François Pitois, Hamida Seba, and Mohammed Haddad*

**Author Index** ..... 399