

12th International Conference on Data Science, Technology and Applications (DATA 2023)

Rome, Italy
11-13 July 2023

Editors:

**Oleg Gusikhin
Slimane Hammoudi
Alfredo Cuzzocrea**

ISBN: 978-1-7138-7646-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda.
at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185

Fax: +351 265520 186

info@scitepress.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

- Data Modeling and AI: From Semantic Networks to Knowledge Graphs 5
Maurizio Lenzerini
- Deep Reinforcement Learning to Improve Traditional Supervised Learning Methodologies 7
Luís Paulo Reis

PAPERS

FULL PAPERS

- Look at the Horizon: Evaluation of a Software Solution Against Cyber Sickness in Virtual Reality Applications 13
Jonathan Harth, Christian-Norbert Zimmer and Michaela Zupanec
- Towards a Low-Code Tool for Developing Data Quality Rules 22
Timon Sebastian Klann, Marcel Altendeitering and Falk Howar
- Benchmarking Automated Machine Learning Methods for Price Forecasting Applications 30
Horst Stühler, Marc-André Zöller, Dennis Klau, Alexandre Beiderwellen-Bedrikow and Christian Tutschku
- Temporal Multidimensional Model for Evolving Graph-Based Data Warehouses 40
Redha Benhissen, Fadila Bentayeb and Omar Boussaid
- Predicting Academic Performance of Low-Income Students in Public Ecuadorian Online Universities: An Educational Data Mining Approach 52
Jorge Rodas-Silva and Jorge Parraga-Alava
- Identifying High-Quality Training Data for Misinformation Detection 64
Jaren Haber, Kornraphop Kawintiranon, Lisa Singh, Alexander Chen, Aidan Pizzo, Anna Pogrebivsky and Joyce Yang
- Hawkes Processes on Social and Mass Media: A Causal Study of the #BlackLivesMatter Movement in the Summer of 2020 77
Alfred Lindström, Simon Lindgren and Raazesh Sainudiin
- Conv-LSTM for Real Time Monitoring of the Mineral Grades in the Flotation Froth 89
Ahmed Bendaouia, El Hassan Abdelwahed, Sara Qassimi, Abdelmalek Boussetta, Intissar Benzakour, Oumkeltoum Amar, François Bourzeix, Achraf Soulala and Oussama Hasidi
- Disease Prediction with Heterogeneous Graph of Electronic Health Records and Toxicogenomics Data 97
Ji-Hyeong Park, Hyun-Soo Choi, Sunhwa Jo and Jinho Kim
- Heterogeneous Ensemble Learning for Modelling Species Distribution: A Case Study of Redstarts Habitat Suitability 105
Omar El Alaoui and Ali Idri

A Proactive Approach for the Sustainable Management of Water Distribution Systems <i>Sarah Di Grande, Mariaelena Berlotti, Salvatore Cavalieri and Roberto Gueli</i>	115
Rigor in Applied Data Science Research Based on DSR: A Literature Review <i>Daniel Szafarski, Tobias Schmieg, Laslo Welz and Thomas Schäffer</i>	126
Extracting Frequent Gradual Patterns Based on SAT <i>Jerry Lonlac, Imen Ouled Dlala, Said Jabbour, Engelbert Mephu Nguifo, Badran Raddaoui and Lakhdar Sais</i>	136
An Advanced BERT_LayerSum Model for Sentiment Classification of COVID-19 Tweets <i>Areeba Umair and Elio Masciari</i>	144
A Comparison Study for Disaster Tweet Classification Using Deep Learning Models <i>Soudabeh Taghian Dinani and Doina Caragea</i>	152
Prediction of QT Prolongation in Advanced Breast Cancer Patients Using Survival Modelling Algorithms <i>Asmir Vodenčarević, Julia Kreuzeder, Achim Wöckel and Peter A. Fasching</i>	164
Anomaly Detection of Medical IoT Traffic Using Machine Learning <i>Lerina Aversano, Mario Luca Bernardi, Marta Cimitile, Debora Montano, Riccardo Pecori and Luca Veltri</i>	173
Semantic, Technical and Legal Interoperability of European Company Open Data in Practice: The STIRData Approach <i>Jakub Klímek, Alexandros Chortaras, Jakub Míšek, Jim J. Yang, Steinar Skagemo and Vassilis Tzouvaras</i>	183
Analyzing Cyber-Physical Systems in Cars: A Case Study <i>Harry H. Beyel, Omar Makke, Fangbo Yuan, Oleg Gusikhin and Wil M.P. van der Aalst</i>	195
SHORT PAPERS	
Structuring the End of the Data Life Cycle <i>Daniel Tebernum and Falk Howar</i>	207
Fundus Unimodal and Late Fusion Multimodal Diabetic Retinopathy Grading <i>Sara El-Ateif and Ali Idri</i>	219
A Visual Analysis of Hazardous Events in Contract Risk Management <i>Georgios Stathis, Giulia Biagioni, Athanasios Trantas, Jaap van den Herik and Bart Custers</i>	227
Pump and Dump Cryptocurrency Detection Using Social Media <i>Domenico Alfano, Roberto Abbruzzese and Domenico Parente</i>	235
Insights from Big Data Economy of Finnish Company Trust, Consumer Confidence and Data Exchange: An Empirical Evidence of Structural Equation Modelling <i>Sunday Adewale Olaleye</i>	241
Embedding-Enhanced Similarity Metrics for Next POI Recommendation <i>Sara Jarrad, Hubert Naacke, Stephane Gancarski and Modou Gueye</i>	247
Behavioral Recommender System for Process Automation Steps <i>Mohammadreza Fani Sani, Fatemeh Nikraftar, Michal Sroka and Andrea Burattin</i>	255

Predicting How Much a Consumer Is Willing to Pay for a Bottle of Wine: Dealing With Data Imbalance <i>Hugo Alonso and Teresa Candeias</i>	263
Exploring Functional Patterns of Driving Records by Interacting with Major Classes and Territory Using Generalized Additive Models <i>Shengkun Xie, Anna T. Lawniczak and Clare Chua-Chow</i>	271
Astronomical Images Quality Assessment with Automated Machine Learning <i>Olivier Parisot, Pierrick Bruneau and Patrik Hitzelberger</i>	279
Building a Dataset for Trip Style Assessment Based on Real Trip Data <i>Luís M. P. Loureiro, Artur J. Ferreira and André R. Lourenço</i>	287
QTrail-DB: A Query Processing Engine for Imperfect Databases with Evolving Qualities <i>Maha Asiri and Mohamed Y. Eltabakh</i>	295
Decomposition Heuristic for the Aircraft Sequencing Problem: Impact on Mathematical and Constraint Programming <i>Joana Leite, Rafael Guedes and Diogo Queirós</i>	303
IntrusionHunter: Detection of Cyber Threats in Big Data <i>Hashem Mohamed, Alia El Bolock and Caroline Sabty</i>	311
Automatic Classification of Quantitative Data from DNS Cache Servers into Stationary and Non-Stationary States Based on Clustering <i>Hikofumi Suzuki and Katsumi Wasaki</i>	319
Characterizing Speed Performance of Multi-Agent Reinforcement Learning <i>Samuel Wiggins, Yuan Meng, Rajgopal Kannan and Viktor Prasanna</i>	327
GUIDO: A Hybrid Approach to Guideline Discovery & Ordering from Natural Language Texts <i>Nils Freyer, Dustin Thewes and Matthias Meinecke</i>	335
Integrating Autoencoder-Based Hybrid Models into Cervical Carcinoma Prediction from Liquid-Based Cytology <i>Ferdous Idrhacen, Ali Idri and Hasnae Zerouaoui</i>	343
Does Categorical Encoding Affect the Interpretability of a Multilayer Perceptron for Breast Cancer Classification? <i>Hajar Hakkoum, Ali Idri, Ibtissam Abnane and José Luis Fernandes-Aleman</i>	351
Enriching Relation Extraction with OpenIE <i>Alessandro Temperoni, Maria Biryukov and Martin Theobald</i>	359
CryptonDL: Encrypted Image Classification Using Deep Learning Models <i>Adham Helbawy, Mahmoud Bahaa and Alia El Bolock</i>	367
A Data Analysis Pipeline for Automating Apple Trait Analysis and Prediction <i>Kyle Ranslam and Ramon Lawrence</i>	375
A Measure Data Catalog for Dashboard Management and Validation <i>Bruno Oliveira, Ana Henriques, Óscar Oliveira, Ana Duarte, Vasco Santos, António Antunes and Elsa Cardoso</i>	381

A Semantic Web Approach for Military Operation Scenarios Development for Simulation <i>André M. Demori, Julio Cesar Cardoso Tesolin, David Fernandes Cruz Moura, João Eduardo Costa Gomes, Gabriel Pedroso, Leonardo Filipe Batista Silva de Carvalho, Edison Pignaton de Freitas and Maria Cláudia Reis Cavalcanti</i>	390
Distributed Edge Computing System for Vehicle Communication <i>Rinith Pakala, Niket Kathiriya, Hossein Haeri, Satya Prasad Maddipatla, Kshitij Jerath, Craig Beal, Sean Brennan and Cindy Chen</i>	398
Comparing Data Store Performance for Full-Text Search: To SQL or to NoSQL? <i>George Fotopoulos, Paris Koloveas, Paraskevi Raftopoulou and Christos Tryfonopoulos</i>	406
Detection of Drowsy Driving Using Wearable Sensors <i>Duarte Pereira, Brigida Monica Faria and Luis Paulo Reis</i>	414
QFLS: A Cloud-Based Framework for Supporting Big Healthcare Data Management and Analytics from Big Data Lakes: Definitions, Requirements, Models and Techniques <i>Alfredo Cuzzocrea and Selim Soufargi</i>	422
Effects of Environmental Conditions on Historic Buildings: Interpretable Versus Accurate Exploratory Data Analysis <i>Marco Parola, Hajar Dirrhami, Mario G. C. A. Cimino and Nunziante Squeglia</i>	429
Detection and Prediction of Leakages in Water Distribution Networks <i>Mariaelena Berlotti, Sarah Di Grande, Salvatore Cavalieri and Roberto Gueli</i>	436
Clustering Object-Centric Event Logs <i>Anahita Farhang Ghahfarokhi, Fatemeh Akoochekian, Fareed Zandkarimi and Wil M. P. van der Aalst</i>	444
Design and Implementation of a Document Encryption Convergence Program Selecting Encryption Methods, and Integrating the Program into the Existing Office System <i>Hong-Jin Ryu and Samuel Sangkon Lee</i>	452
The Imbalance Data Handling of XGBoost in Insurance Fraud Detection <i>Nathanael Theovanny Averro, Hendri Murfi and Gianinna Ardaneswari</i>	460
BERT-Based Hybrid Deep Learning with Text Augmentation for Sentiment Analysis of Indonesian Hotel Reviews <i>Maxwell Thomson, Hendri Murfi and Gianinna Ardaneswari</i>	468
Optimization of Surgery Scheduling Problems Based on Prescriptive Analytics <i>João Lopes, Gonçalo Vieira, Rita Veloso, Susana Ferreira, Maria Salazar and Manuel Filipe Santos</i>	474
An Ontology-Based Collaborative Business Intelligence Framework <i>Muhammad Fahad and Jérôme Darmont</i>	480
Creation and Evaluation of a Food Product Image Dataset for Product Property Extraction <i>Christoph Brosch, Alexander Bouwens, Sebastian Bast, Swen Haab and Rolf Krieger</i>	488
Geo-Semantic Event-POI Matching of Large Mobility Datasets <i>Ndiouma Bame, Ibrahima Gueye and Hubert Naacke</i>	496
Detecting Anomalies on Cryptocurrency Markets Using Graph Algorithms <i>Agata Skorupka</i>	504

Behavioral Modeling of Real Dynamic Processes in an Industry 4.0-Oriented Context <i>Dylan Molinié, Kurosh Madani and Véronique Amarger</i>	510
Dense Information Retrieval on a Latin Digital Library via LaBSE and LatinBERT Embeddings <i>Federico Andrea Galatolo, Gabriele Martino, Mario G. C. A. Cimino and Chiara Ombretta Tommasi</i>	518
Performance Evaluation and Comparison of a New Regression Algorithm <i>Sabina Gooljar, Kris Manohar and Patrick Hosein</i>	524
A Multi-Factor Approach to Measure User Preference Similarity in Neighbor-Based Recommender Systems <i>Ho Thi Hoang Vy, Tiet Gia Hong, Vu Thi My Hang, Cuong Pham-Nguyen and Le Nguyen Hoai Nam</i>	532
Union k-Fold Feature Selection on Microarray Data <i>Artur J. Ferreira and Mário A. T. Figueiredo</i>	540
Towards Data Ecosystems in Smart Cities: Success Factors of Urban Data Spaces <i>Josh Haberkern and Thomas Schäffer</i>	548
Decision Support System for Corporate Reputation Based Social Media Listening Using a Cross-Source Sentiment Analysis Engine <i>R. E. Loke and S. Pathak</i>	559
EREO: An Effective Rule Evaluation Framework for Discovering Interesting Patterns in US Birth Data and Beyond <i>Abhilash C. B. and Kavi Mahesh</i>	568
Sharding and Master-Slave Replication of NoSQL Databases: Comparison of MongoDB and Redis <i>Anikó Vágner and Mustafa Al-Zaidi</i>	576
A Design Framework for a Blockchain-Based Open Market Platform of Enriched Card-Based Transactional Data for Big Data Analytics and Open Banking <i>Trevor Toy and Josef Langerman</i>	583
The Application of Affective Measures in Text-Based Emotion Aware Recommender Systems <i>John Kalung Leung, Igor Griva, William G. Kennedy, Jason M. Kinser, Sohyun Park and Seo Young Lee</i>	590
AutoImpute: An Autonomous Web Tool for Data Imputation Based on Extremely Randomized Trees <i>Mustafa Alabadla, Fatimah Sidi, Iskandar Ishak, Hamidah Ibrahim, Hazlina Hamdan, Shahril Iskandar Amir and Appak Yessirkep Nurlankyzy</i>	598
On Data-Preprocessing for Effective Predictive Maintenance on Multi-Purpose Machines <i>Lukas Meitz, Michael Heider, Thorsten Schöler and Jörg Hähner</i>	606
Optimized Fusion Chain with Web3 Storage <i>Tahiya Tabassum Lisa and Chaiyachet Saivichit</i>	613
GreenCC: A Hybrid Approach to Sustainably Validate Manufacturing Data in Industry 4.0 Environments <i>Simon Paasche and Sven Groppe</i>	621
A Successive Quadratic Approximation Approach for Tuning Parameters in a Previously Proposed Regression Algorithm <i>Patrick Hosein, Kris Manohar and Ken Manohar</i>	629
AUTHOR INDEX	635