13th International Conference on Data Science, Technology and Applications (DATA 2024)

Dijon, France
9-11 July 2024

Editors:

Alfredo Cuzzocrea
Elhadj Benkhelifa
Oleg Gusikhin
Slimane Hammoudi

ISBN: 979-8-3313-0548-2

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© (2024) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

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

Towards a Responsible Thinking in the New Era of Gen AI: Walking the Walk 5
Mona Diab

Scalable Vector Analytics: A Story of Twists and Turns 7
Themis Palpanas

Past, Present, Future, and Far Future of AI 9
Jürgen Schmidhuber

PAPERS

FULL PAPERS

Artificial Intelligence-Powered Large Language Transformer Models for Opioid Abuse and Social Determinants of Health Detection for the Underserved Population 15
Don Roosan, Yanting Wu, Jay Chok, Christopher P. Sanine, Tiffany Khou, Yawen Li and Hasiba M. Khan

Online Machine Learning for Adaptive Ballast Water Management 27
Nadeem Iftikhar, Yi-Chen Lin, Xiufeng Liu and Finn Ebertsen Nordbjerg

Efficient and Secure Multiparty Querying over Federated Graph Databases 39
Nouf Aljuaid, Alexei Lisitsa and Sven Schewe

Reputation, Sentiment, Time Series and Prediction 51
Peter Mitic

A Cascade of Consequences: Improving an Accident Analysis Method by Learning from a Real Life Telecommunications Accident 62
Hans C. A. Wienen, Faiza A. Bukhsh, Eelco Vriezekolk and Luís Ferreira Pires

Is Positive Sentiment Missing in Corporate Reputation? 71
Peter Mitic

A Deep Dive into GPT-4's Data Mining Capabilities for Free-Text Spine Radiology Reports 82
Klaudia Szabó Ledenyi, András Kicsi and László Vidács

Privacy-Preserving Big Hierarchical Data Analytics via Co-Occurrence Analysis 93
Alfredo Cuzzocrea and Selim Soufargi

Risk-Stratified Multi-Objective Resource Allocation for Optimal Aviation Security 104
Eva K. Lee, Taylor J. Leonard and Jerry C. Booker

Cross-Lingual Low-Resources Speech Emotion Recognition with Domain Adaptive Transfer Learning 118
Imen Baklouti, Olfa Ben Ahmed and Christine Fernandez-Maloigne

HTC-GEN: A Generative LLM-Based Approach to Handle Data Scarcity in Hierarchical Text Classification 129
Carmelo Fabio Longo, Misael Mongiovı, Luana Bulla and Giusy Giulia Tuccari

Course Recommendation System for Company Job Placement Using Collaborative Filtering and Hybrid Model 139
Jaeheon Park, Suan Lee, Woncheol Lee and Jinho Kim

Intelligent Sampling System for Connected Vehicle Big Data 150
Omar Makke, Syam Chand, Vamsee Krishna Batchu, Oleg Gusikhin and Vicky Svidenko

An Exploratory Analysis of Malaria and Climatic Factors in India 159
Sachin Yadav Bodke and Usha Ananthakumar

Intelligent Transportation Systems: A Survey on Data Engineering 169
Safa Batita, Achraf Makni and Ikram Amous

Expert Agent Guided Learning with Transformers and Knowledge Graphs 180
Chukwuka Victor Obionwu, Bhavya Baburaj Chovatta Valappil, Minu Genty, Maria Jomy, Visakh Padmanabhan, Aishwarya Suresh, Sumat Singh Bedi, David Broneske and Gunter Saake

A Max Flow Min Cut View of Social Media Posts 190
James Abello, Timothy R. Tangherlini and Haoyang Zhang

SHORT PAPERS

Brain Stroke Prediction Using Visual Geometry Group Model 205
V. Narayanan, A. Reddy, V. Venkatesh, S. Tutun, P. Norouzzadeh, E. Snir, S. Mahmoud and B. Rahmani

Stochastic Simulation Agent for Unknown Inventory Demands in Healthcare Supply Management 211
Rafael Marin Machado de Souza, Leandro Nunes de Castro, Marcio Biczyk, Marcos dos Santos and Eder Costa Cassettari

Dynamic Price Prediction for Revenue Management System in Hospitality Sector 218
Susanna Saitta, Vito D'Amico and Giovanni Maria Farinella

An Effective Prediction of Events in Social Networks Using Influence Score of Communities 229
B. S. A. S. Rajita, Yaganti Bhanu Vikas, Pritish Prashant Moharir, Deepa Kumari and Subhrakanta Panda

Logical Rule Set to Data Acquisition and Database Semantics 237
Susumu Yamasaki and Mariko Sasakura

A Data-Driven Approach for Predictive Maintenance of Impellers in Flexible Impeller Pumps Using Prophet 245
Efe Can Demir and Sencer Sultanoğlu

Objective Evaluation of Sleep Disturbances in Older Adults with Cognitive Impairment Using a Bed Sensor System and Self-Organizing Map Analysis 253
Tomoko Kamimura, Risa Otsuka, Asaka Domoto, Hikofumi Suzuki and Mamino Tokita

BeRTo: An Efficient Spark-Based Tool for Linking Business Registries in Big Data Environments 259
Andrea Colombo and Francesco Invernici

Recommendations of Research Articles by Experts: Visualizing Relationships and Expertise 269
Peiling Wang, Scott Shumate, Pinghao Ye and Chad Mitchell

Data-Driven Model Categorization: Advancing Physical Systems Analysis Through Graph Neural Networks 277
Andrija Grbavac, Martin Angerbauer, Michael Grill and André Casal Kulzer

Students' Performance in Learning Management System: An Approach to Key Attributes Identification and Predictive Algorithm Design 285
Dynil Duch, Madeth May and Sébastien George

Dataset Balancing in Disease Prediction 293
Vincenza Carchiolo and Michele Malgeri

A Comparison of the Efficiencies of Various Structured and Semi- Structured Data Formats in Data Analysis and Big Data Analytic Development 301
Heather E. Graham and Taoxin Peng

Forgetting in Knowledge Graph Based Recommender Systems 309
Xu Wang and Christopher Brewster

Towards Computational Performance Engineering for Unsupervised Concept Drift Detection: Complexities, Benchmarking, Performance Analysis 318
Elias Werner, Nishant Kumar, Matthias Lieber, Sunna Torge, Stefan Gumhold and Wolfgang E. Nagel

Influential Factors on Drivetrain Consumption in Electric City Buses and Assessing the Optimization Potentials 330
Sunilkumar Raghuraman, Daniel Baumann, Marc Schindewolf and Eric Sax

Yet Another Miner Utility Unveiling a Dataset: CodeGrain 338
Dániel Horváth and László Vidács

Use of Semantic Artefacts in Agricultural Data-Driven Service Development 346
Silke Cuno and Philipp Lämmel

Study of Impact of Gender on Engagement and Performance of Engineering Students 358
M. E. Sousa-Vieira, J. C. López-Ardao and M. Fernández-Veiga

Random Neural Network Ensemble for Very High Dimensional Datasets 368
Jesus S. Aguilar–Ruiz and Matteo Fratini

Federated Road Surface Anomaly Detection Using Smartphone Accelerometer Data 376
Oussama Mazari Abdessameud and Walid Cherifi

A Comparative Study on the Impact of Categorical Encoding on Black Box Model Interpretability 384
Hajar Hakkoum and Ali Idri

Comparative Analysis of Hate Speech Detection Models on Brazilian Portuguese Data: Modified BERT vs. BERT vs. Standard Machine Learning Algorithms 392
Thiago Mei Chu, Leila Weitzel and Paulo Quaresma

Real-Time Equipment Health Monitoring Using Unsupervised Learning Techniques 401
Nadeem Iftikhar and Finn Ebertsen Nordbjerg

Machine Learning Classification in Cardiology: A Systematic Mapping Study 409
Khadija Anejjar, Fatima Azzahra Amazal and Ali Idri

Impact of Satellites Streaks for Observational Astronomy: A Study on Data Captured During One Year from Luxembourg Greater Region 417
Olivier Parisot and Mahmoud Jaziri

Strategic Placement of Data Centers for Economic Analysis: An Online Algorithm Approach 425
Christine Markarian and Claude Fachkha

Discretization Strategies for Improved Health State Labeling in Multivariable Predictive Maintenance Systems 434
Jean-Victor Autran, Véronique Kuhn, Jean-Philippe Diguet, Matthias Dubois and Cédric Buche

Failure Prediction Using Multimodal Classification of PCB Images 442
Pedro M. Goncalves, Miguel A. Brito and Jose Guilherme Cruz Moreira

A Few-Shot Learning-Focused Survey on Recent Named Entity Recognition and Relation Classification Models 450
Sakher Khalil Alqaaidi, Elika Bozorgi, Afsaneh Shams and Krzysztof J. Kochut

It is Time to Develop an Auditing Framework to Promote Value Aware Chatbots 460
Yanchen Wang and Lisa Singh

Towards FAIR Data Workflows for Multidisciplinary Science: Ongoing Endeavors and Future Perspectives in Plasma Technology 471
Robert Wagner, Dagmar Waltemath, Kristina Yordanova and Markus M. Becker

A Web-Based Hate Speech Detection System for Dialectal Arabic 478
Anis Charfi, Andria Atalla, Raghda Akasheh, Mabrouka Bessghaier and Wajdi Zaghouani

Interoperable Open Data Platforms: A Prototype for Sharing CKAN Data Sources 486
Sebastian Becker and Marcel Altendeitering

Twitter Metrical Data Analysis Using R: Twiplomacy in the Outbreak of the War in Ukraine 494
Dimitrios Vagianos and Thomas Papatsas

Anomaly Detection in Industrial Production Products Using OPC-UA and Deep Learning 505
Henry O. Velesaca, Doménica Carrasco, Dario Carpio, Juan A. Holgado-Terriza, Jose M. Gutierrez-Guerrero, Tonny Toscano and Angel D. Sappa

Prediction of Academic Success in a University and Improvement Using Lean Tools 513
Kléber Sánchez and Diego Vallejo-Huanga

Machine Learning for KPI Development in Public Administration 522
Simona Fioretto, Elio Masciari and Enea Vincenzo Napolitano

Visualization and Interpretation of Mel-Frequency Cepstral Coefficients for UAV Drone Audio Data 528
Mia Y. Wang, Zhiwei Chu, Conner Entzminger, Yi Ding and Qian Zhang

Spatial and Spatio-Temporal Modelling of Auto Insurance Claim Frequencies During Pre-and Post-COVID-19 Pandemic 535
Jin Zhang, Shengkun Xie, Anna T. Lawniczak and Clare Chua-Chow

Towards Semantic Data Management Plans for Efficient Review Processing and Automation 543
Jana Martínková and Marek Suchánek

Integration and Optimization of XNAT-Based Platforms for the Management of Heterogeneous and Multicenter Data in Biomedical Research 551
Camilla Scapicchio, Silvia Arezzini, Maria Evelina Fantacci, Antonino Formuso, Aafke Christine Kraan, Enrico Mazzoni, Sara Saponaro, Maria Irene Tenerani and Alessandra Retico

Design Features for Data Trustee Selection in Data Spaces 559
Michael Steinert, Daniel Tebernum and Marius Hupperz

Visualizing OWL and RDF: Advancing Ontology Representation for Enhanced Semantic Clarity and Communication 571
Giulia Biagioni

Navigating the AI Timeline: From 1995 to Today 577
Vincenza Carchiolo and Michele Malgeri

Random Forest Classification of Cognitive Impairment Using Digital Tree Drawing Test (dTDT) Data 585
Sebastian Unger, Zafer Bayram, Laura Anderle and Thomas Ostermann

Presence of Corporate Reputation Cues in Company Vacancy Texts Boosts Vacancy Attractiveness as Perceived by Employees 593
R. E. Loke and F. Betten

AUTHOR INDEX 605