

11th International Conference on Data Science, Technology and Applications (DATA 2022)

Lisbon, Portugal
11 – 13 July 2022

Editors:

**Alfredo Cuzzocrea
Oleg Gusikhin**

**Wil van der Aalst
Slimane Hammoudi**

ISBN: 978-1-7138-6386-1

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© (2022) 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

From Supply Chain Resilience to Ecosystem Viability: Lessons from and Thinking Beyond the COVID-19 Pandemic 5
Dmitry Ivanov

The Difficult Path to Become Datadriven 7
Lionel Rigaud

PAPERS

FULL PAPERS

Development of a Semantic Database Model to Facilitate Data Analytics in Battery Cell Manufacturing 13
Ozan Yesilyurt, David Brandt, Julian Joël Grimm, Kamal Hussein, Aleksandra Naumann, Julia Meiners and David Becker-Koch

Startable: Multidimensional Modelling for Column-Oriented NoSQL 21
Leandro Mendes Ferreira, Solange Nice Alves-Souza and Luciana Maria da Silva

Different Metrics Results in Text Summarization Approaches 31
Marcello Barbella, Michele Risi, Genoveffa Tortora and Alessia Auriemma Citarella

Preprocessing of Terrain Data in the Cloud using a Workflow Management System 40
Marvin Kaster and Hendrik M. Würz

Detecting Bots in Social-networks using Node and Structural Embeddings 50
Ashkan Dehghan, Kinga Siuta, Agata Skorupka, Akshat Dubey, Andrei Betlen, David Miller, Wei Xu, Bogumił Kamiński and Paweł Pralat

Highly Automated Corner Cases Extraction: Using Gradient Boost Quantile Regression for AI Quality Assurance 62
Niels Heller and Namrata Gurung

Development of a Text Classification Framework using Transformer-based Embeddings 74
Sumona Yeasmin, Nazia Afrin, Kashfia Saif and Mohammad Rezwanaul Huq

Exploratory Analysis of Chat-based Black Market Profiles with Natural Language Processing 83
André Büsgen, Lars Klöser, Philipp Kohl, Oliver Schmidts, Bodo Kraft and Albert Zündorf

Domain-independent Data-to-Text Generation for Open Data 95
Andreas Burgdorf, Micaela Barkmann, André Pomp and Tobias Meisen

Inferring #MeToo Experience Tweets using Classic and Neural Models 107
Julianne Zech, Lisa Singh, Kornraphop Kawintiranon, Naomi Mezey and Jamillah Bowman Williams

A Formal Model to Support Discourse Semantic Landscape Analysis 118
Isabelle Linden, Bruno Dumas and Anne Wallemacq

A Model based on 2-tuple Linguistic Model and CRITIC Method for Hotel Classification <i>Ziwei Shu, Ramón Alberto Carrasco González, Javier Portela García-Miguel and Manuel Sánchez-Montañés</i>	127
A2P: @ <u>attention</u> -based Memory @ <u>access</u> @ <u>prediction</u> for Graph Analytics <i>Pengmiao Zhang, Rajgopal Kannan, Anant V. Nori and Viktor K. Prasanna</i>	135
Tackling Model Drifts in Industrial Model-driven Software Product Lines by Means of a Graph Database <i>Christof Tinnes, Uwe Hohenstein, Wolfgang Rössler and Andreas Biesdorf</i>	146
Big Data Analysis of Ionosphere Disturbances using Deep Autoencoder and Dense Network <i>Rayan Abri, Harun Artuner, Sara Abri and Salih Cetin</i>	158
Enhancing Biomedical Scientific Reviews Summarization with Graph-based Factual Evidence Extracted from Papers <i>Giacomo Frisoni, Paolo Italiani, Francesco Boschi and Gianluca Moro</i>	168
Modelling of Efficient Graph-aware Data Storage using DNA <i>Asad Usmani and Lena Wiese</i>	180
SHORT PAPERS	
Using the Silhouette Coefficient for Representative Search of Team Tactics in Noisy Data <i>Friedemann Schwenkreis</i>	193
Position Paper: Quality Assurance in Deep Learning Systems <i>Domingos F. Oliveira and Miguel A. Brito</i>	203
Supporting Trainset Annotation for Text Classification of Incoming Enterprise Documents <i>Juris Rats and Inguna Pedre</i>	211
From Digital Footprints to Facts: Mining Marketing Policies of the Greek Community on Instagram and Youtube <i>Dimitris Linarakis, Stefanos Vlachos and Paraskevi Raftopoulou</i>	219
User Profiling: On the Road from URLs to Semantic Features <i>Claudio Barros and Perrine Moreau</i>	227
Maneuver-based Visualization of Similarities between Recorded Traffic Scenarios <i>Thilo Braun, Lennart Ries, Moritz Hesche, Stefan Otten and Eric Sax</i>	236
The Role of Fake Review Detection in Managing Online Corporate Reputation <i>R. E. Loke and Z. Kisoen</i>	245
Towards Decentralized Parameter Servers for Secure Federated Learning <i>Muhammad El-Hindi, Zheguang Zhao and Carsten Binnig</i>	257
Sensitivity Analysis using Regression Models: To Determine the Impact of Meta-level Features on the Youtube Views <i>Vaishnavi Borwankar, Catherine Chris, Hitesh Kumar and Sophia Rahaman</i>	270
Evaluation by Simulation of the Diffusion Methods in the Cloud: Based Network Architecture for Digital Open Universities <i>Boukar Abatchia Nicolas, Mahamadou Issoufou Tiado, Nassirou Adamou Hassane and Ibrahim Ganaou Noura</i>	276

Framework for Public Health Policy Indicators Governance and Metadata Quality Flags to Promote Data Literacy <i>Wesley Lourenco Barbosa, Jacqueline David Planas, Maritza M. Carvalho Francisco, Solange Nice Alves-Souza, Lucia Vilela Leite Filgueiras, Leandro Manuel Reis Velloso and Luiz Sérgio de-Souza</i>	281
A Hybrid Approach for Product Classification based on Image and Text Matching <i>Sebastian Bast, Christoph Brosch and Rolf Krieger</i>	293
Compiling Open Datasets in Context of Large Organizations while Protecting User Privacy and Guaranteeing Plausible Deniability <i>Igor Jakovljevic, Christian Gütl, Andreas Wagner and Alexander Nussbaumer</i>	301
A Comparison of Automatic Labelling Approaches for Sentiment Analysis <i>Sumana Biswas, Karen Young and Josephine Griffith</i>	312
Feature Extraction and Failure Detection Pipeline Applied to Log-based and Production Data <i>Rosaria Rossini, Nicolò Bertozzi, Eliseu Pereira, Claudio Pastrone and Gil Gonçalves</i>	320
Multitudinous Data Platform for Community Big Data <i>S. Junrat, J. Nopparat, M. Manopiroonporn, W. Suntiarnornut and S. Charoenpanyasak</i>	328
Towards Semantic Interoperability of Core Registers in Croatia <i>Kornelije Rabuzin, Darko Gulija, Leo Mršić and Nikola Modrušan</i>	337
Unsupervised Electrodermal Data Analysis Comparison between Biopac and Empatica E4 Data Collection Platforms <i>Kassy Raymond and Andrew Hamilton-Wright</i>	345
Analysis of Data Quality in Digital Smart Cities: The Cases of Nantes, Hamburg and Helsinki <i>José L. Hernández, Ana Quijano, Rubén García, Pierre Nouaille, Lukas Risch, Mikko Virtanen and Ignacio de Miguel</i>	353
Towards Indoor Radon Analytics: An OLAP-based Multidimensional Approach <i>Rolando Azevedo, Joaquim P. Silva, Nuno Lopes, António Curado, Leonel J. R. Nunes and Sérgio Ivan Lopes</i>	361
Analyzing Cross-impact Matrices for Managerial Decision-making Problems with the DEMATEL Approach <i>Shailesh Tripathi, Nadine Bachmann, Manuel Brunner and Herbert Jodlbauer</i>	370
Estimating the Optimal Number of Clusters from Subsets of Ensembles <i>Afees Adegoke Odebode, Allan Tucker, Mahir Arzoky and Stephen Swift</i>	383
End-to-End Data Quality: Insights from Two Case Studies <i>M. Redwan Hasan and Christine Legner</i>	392
Prediction of Sulfur in the Hot Metal based on Data Mining and Artificial Neural Networks <i>Wanderleiton Cardoso and Rendo di Felice</i>	400
Students or Mechanical Turk: Who Are the More Reliable Social Media Data Labelers? <i>Lisa Singh, Rebecca Vanarsdall, Yanchen Wang and Carole Roan Gresenz</i>	408
Using Machine Learning Methods and the Influenza Simulation System to Explore the Similarities of Taiwan's Administrative Regions <i>Zong-Kai Lai, Yi-Ting Chiang, Tsan-sheng Hsu and Hung-Jui Chang</i>	416

Exploring Corporate Reputation based on Sentiment Polarities That Are Related to Opinions in Dutch Online Reviews <i>R. E. Loke and J. Vergeer</i>	423
Improved Boosted Classification to Mitigate the Ethnicity and Age Group Unfairness <i>Ivona Colakovic and Sašo Karakatič</i>	432
Working Efficiently with Large Geodata Files using Ad-hoc Queries <i>Pascal Bormann, Michel Krämer and Hendrik M. Würz</i>	438
Interoperability-oriented Quality Assessment for Czech Open Data <i>Dasa Kusnirakova, Mouzhi Ge, Leonard Wallezky and Barbora Buhnova</i>	446
Political Analytics on Election Candidates and Their Parties in Context of the US Presidential Elections 2020 <i>Kalpdrum Passi and Rakshit Sorathiya</i>	454
Automated Neoclassical Vertical Canon Validation in Human Faces with Machine Learning <i>Ashwinee Mehta, Maged Abdelaal, Moamen Sheba and Nic Herndon</i>	461
Towards Programmable Memory Controller for Tensor Decomposition <i>Sasindu Wijeratne, Ta-Yang Wang, Rajgopal Kannan and Viktor Prasanna</i>	468
Evaluation of Architectures for FAIR Data Management in a Research Data Management Use Case <i>Benedikt Heinrichs, Marius Politze and M. Amin Yazdi</i>	476
A Two-tire Approach for Organization Name Entity Resolution <i>Almuth Müller and Achim Kuwertz</i>	484
Line-up Optimization Model of Basketball Players and the Simulation Evaluation <i>Wang Yichen and Haruka Yamashita</i>	492
Using Convolutional Neural Networks for Detecting Acrylamide in Biscuit Manufacturing Process <i>Dilruba Topcuoglu, Berat Utkan Mentés, Nur Askin, Ayse Damla Sengul, Zeynep Deniz Cankut, Talip Akdemir, Murat Ayvaz, Elif Kurt, Ozge Erdohan, Tumay Temiz and Murat Ceylan</i>	500
Auctions and Estimates: Evidence from Indian Art Market <i>Shailendra Gurjar and Usha Ananthakumar</i>	504
Clustered Majority Judgement <i>Emanuele d'Ajello, Davide Formica, Elio Masciari, Gaia Mattia, Arianna Anniciello, Cristina Moscariello, Stefano Quintarelli and Davide Zaccarella</i>	512
Local Personal Data Processing with Third Party Code and Bounded Leakage <i>Robin Carpentier, Iulian Sandu Popa and Nicolas Anceaix</i>	520
Functional Component Descriptions for Electrical Circuits based on Semantic Technology Reasoning <i>Johannes Bayer, Mina Karami Zadeh, Markus Schröder and Andreas Dengel</i>	528
Efficient Subgraph Indexing for Biochemical Graphs <i>Chimi Wangmo and Lena Wiese</i>	533
Slide-recommendation System: A Strategy for Integrating Instructional Feedback into Online Exercise Sessions <i>Victor Obionwu, Vincent Toulouse, David Broneske and Gunter Saake</i>	541

ADiBA Big Data Adoption Framework: Accelerating Big Data Revolution 5.0 <i>Norhayati Daut, Naomie Salim, Chan Weng Howe, Anazida Zainal, Sharin Hazlin Huspi, Masitah Ghazali and Fatimah Shafinaz Ahmad</i>	549
NoSQL Document Databases Assessment: Couchbase, CouchDB, and MongoDB <i>Inês Carvalho, Filipe Sá and Jorge Bernardino</i>	557
AUTHOR INDEX	565