CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

On the Pareto Principle in Process Mining, Task Mining, and Robotic Process Automation
Wil M. P. van der Aalst

From Data to the Press: Data Management for Journalism and Fact-Checking
Ioana Manolescu

PAPERS

FULL PAPERS

Auxiliary Decision-making for Controlled Experiments based on Mid-term Treatment Effect Prediction: Applications in Ant Financial’s Offline-payment Business
Gang Li and Huizhi Xie

A Framework for Creating Policy-agnostic Programming Languages
Fabian Bruckner, Julia Pampus and Falk Howar

Capability-based Scheduling of Scientific Workflows in the Cloud
Michel Krämer

Integrating Lightweight Compression Capabilities into Apache Arrow
Juliana Hildebrandt, Dirk Habich and Wolfgang Lehner

Classification of Products in Retail using Partially Abbreviated Product Names Only
Oliver Allweyer, Christian Schorr, Rolf Krieger and Andreas Mohr

On Generating Efficient Data Summaries for Logistic Regression: A Coreset-based Approach
Nery Riquelme-Granada, Khuong An Nguyen and Zhiyuan Luo

Catalog Integration of Low-quality Product Data by Attribute Label Ranking
Oliver Schmidts, Bodo Kraft, Marvin Winkens and Albert Zündorf

A Gradient Descent based Heuristic for Solving Regression Clustering Problems
Enis Kayış

Is Open Data Ready for Use by Enterprises? Learnings from Corporate Registers
Pavel Krasikov, Timo Obrecht, Christine Legner and Markus Eurich

Learning Interpretable and Statistically Significant Knowledge from Unlabeled Corpora of Social Text Messages: A Novel Methodology of Descriptive Text Mining
Giacomo Frisoni, Gianluca Moro and Antonella Carbonaro
SHORT PAPERS

Applied Data Science: An Approach to Explain a Complex Team Ball Game
Friedemann Schwenkreis and Eckard Nothdurft

Trading Desk Behavior Modeling via LSTM for Rogue Trading Fraud Detection
Marine Neyret, Jaouad Ouaggag and Cédric Allain

Leveraging Machine Learning for Fake News Detection
Elio Masciari, Vincenzo Moscato, Antonio Picariello and Giancarlo Sperli

Multi-view Clustering Analyses for District Heating Substations
Shahrooz Abghari, Veselka Boeva, Jens Brage and Håkan Grahn

Automatic Detection of Gait Asymmetry
Maciej Cwierlikowski and Mercedes Torres Torres

Hybrid Multi-model Multi-platform (HM3P) Databases
Sven Groppe and Jinghua Groppe

Improving Statistical Reporting Data Explainability via Principal Component Analysis
Shengkun Xie and Clare Chua-Chow

Toward a New Quality Measurement Model for Big Data
Mandana Omidbakhsh and Olga Ormandjieva

Improving Public Sector Efficiency using Advanced Text Mining in the Procurement Process
Nikola Modrušan, Kornelije Rabuzin and Leo Mršić

Reference Data Abstraction and Causal Relation based on Algebraic Expressions
Susumu Yamasaki and Mariko Sasakura

Real-time Visualization of Sensor Data in Smart Manufacturing using Lambda Architecture
Nadeem Iftikhar, Bartosz Piotr Lachowicz, Akos Madaras, Finn Ebertsen Nordbjerg, Thorkil Baattrup-Andersen and Karsten Jeppesen

A Conceptual Framework for a Flexible Data Analytics Network
Daniel Tebernum and Dustin Chabrowski

Identification of Social Influence on Social Networks and Its Use in Recommender Systems: A Systematic Review
Lesly Alejandra Gonzalez Camacho and Solange Nice Alves-Souza

Trust Profile based Trust Negotiation for the FHIR Standard
Eugene Sanzi and Steven A. Demurjian

iTLM: A Privacy Friendly Crowdsourcing Architecture for Intelligent Traffic Light Management
Christian Roth, Mirja Nitschke, Matthias Hörmann and Doğan Kesdoğan

Initializing k-means Clustering
Christian Borgelt and Olha Yarikova

The Need for an Enterprise Risk Management Framework for Big Data Science Projects
Jeffrey Saltz and Sucheta Lahiri
Towards Large-scale Gaussian Process Models for Efficient Bayesian Machine Learning

Fabian Berns and Christian Beecks

Ensemble Clustering based Semi-supervised Learning for Revenue Accounting Workflow Management

Tianshu Yang, Nicolas Pasquier and Frederic Precioso

Predicting the Environment of a Neighborhood: A Use Case for France

Nelly Barret, Fabien Duchateau, Franck Favetta and Loïc Bonneval

Context-aware Retrieval and Classification: Design and Benefits

Kurt Englmeier

Sentiment Polarity Classification of Corporate Review Data with a Bidirectional Long-Short Term Memory (biLSTM) Neural Network Architecture

R. E. Loke and O. Kachaniuk

Towards Self-adaptive Defect Classification in Industrial Monitoring

Andreas Margraf, Jörg Höhner, Philipp Bramt and Steffen Geinitz

A CART-based Genetic Algorithm for Constructing Higher Accuracy Decision Trees

Elif Ersoy, Eringc Albey and Enis Kayış

AUTHOR INDEX