DATA ANALYTICS 2021

The Tenth International Conference on Data Analytics

October 3 - 7, 2021

Barcelona, Spain

DATA ANALYTICS 2021 Editors

Sandjai Bhulai, Vrije Universiteit Amsterdam, The Netherlands

Ivana Semanjski, Ghent University, Belgium

Les Sztandera, Thomas Jefferson University, USA

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© (2021) by International Academy, Research, and Industry Association (IARIA)
Please refer to the Copyright Information page.

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

International Academy, Research, and Industry Association (IARIA)
412 Derby Way
Wilmington, DE 19810

Phone: (408) 893-6407
Fax: (408) 527-6351

petre@iaria.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

Table of Contents

A Comparison of Machine-Learned Survival Models for Predicting Tenure from Unstructured Resumes 1
Corne de Ruijt, Vladimer Kobayashi, and Sandjai Bhulai

Electric Vehicle Charging Locations for Uber in New York City 7
Victor Salazar Ramos, Andreea Mateescu, Elisabeth Fokker, Jacky P.K. Li, and Sandjai Bhulai

Feature Engineering and Machine Learning Modelling for Predictive Maintenance Based on Production and Stop Events 14
Ariel Cedola, Rosaria Rossini, Ilaria Bosi, and Davide Conzon

Analyzing the United State's Nationwide Opioid Crisis and Socio-economic Factors using K-Means Clustering 23
Ryan McGinnis and Les Sztandera

Data Merging Technique in Cataract Patients in Telangana for Enhancing Public Awareness of Visual Impairment 28
Amna Alalawi and Les Sztandera

How Do Socioeconomic Factors Correlate to COVID-19 Cases and Deaths? 33
Anthony Guzman and Yoo Jin

Integrated Architecture of SQL Engine and Data Analytics Tool with Apache Arrow Flight and Its Performance Evaluation 40
Yuichiro Aoki and Satoru Watanabe

Analysis of Minimal Clearance and Algorithm Selection Effect on Path Planning for Autonomous Systems 45
Ronald Ponguillo-Intriago, Payam Khazaelpour, Ignacio Querol Puchal, Silvio Semanjski, Daniel Ochoa, Sidharta Gautama, and Ivana Semanjski

Detection of Concept Drift in Manufacturing Data with SHAP Values to Improve Error Prediction 51
Christian Seiffer, Holger Ziekow, Ulf Schreier, and Alexander Gerling

Discovering DataOps: A Comprehensive Review of Definitions, Use Cases, and Tools 61
Kiran Mainali, Lisa Ehrlinger, Johannes Himmelbauer, and Mihhail Matskin

Feature Engineering vs Feature Selection vs Hyperparameter Optimization in the Spotify Song Popularity Dataset 70
Alan Cueva Mora and Brendan Tierney

A Concept for a Comprehensive Understanding of Communication in Mobile Forensics 74
Jian Xi, Michael Spranger, and Dirk Labudde

Classification of Bots and Gender using Topic Unigrams 77
Astrid Fleig, Lisa Geyersbach, Melissa Gohler, Patricia Kurz, Paul Limburg, Dirk Labudde, and Michael

Spranger

Evaluation of Filter Methods for Feature Selection by Using Real Manufacturing Data 82
Alexander Gerling, Holger Ziekow, Ulf Schreier, Christian Seiffer, Andreas Hess, and Djaffar Abdeslam

Modelling the Consistency between Customer Opinion and Online Rating with VADER Sentiment and Bayesian Networks 92
Alexandros Bousdekis, Dimitris Kardaras, and Stavroula Barbounaki