

Petra Perner (Ed.)

Advances in Data Mining

Applications and Theoretical Aspects

21th Industrial Conference, ICDM 2021
New York, USA, July 14 – 18 2021
Proceedings

Volume Editor

Prof. Dr. Petra Perner
FutureLab Artificial Intelligence IBaI-2

PF 10128
01446 Radeberg
E-mail: pperner@futurelab-ai-ibai-2.de

The German National Library listed this publication in the
German National Bibliography.
Detailed bibliographical data can be downloaded from [http://
dnb.ddb.de](http://dnb.ddb.de).

ibai-publishing
Prof. Dr. Petra Perner
PF 30 11 38
04251 Leipzig, Germany
E-mail: info@ibai-publishing.org
<http://www.ibai-publishing.org>

P-ISSN 1864-9734
E-ISSN 2699-5220
ISBN 978-3-942952-82-8

Copyright © 2021 ibai-publishing

Table of Content

BERTSurv: BERT-Based Survival Models for Predicting Outcomes of Trauma Patients <i>Yun Zhao, Qinghang Hong, Xinlu Zhang, Yu Deng, Yuqing Wang, Linda Petzold</i>	1
Method of Computing Similarity with Clusters <i>Arthur Yosef, Moti Schneider</i>	17
Empirical Analysis of Machine Learning Configurations for Prediction of Multiple Organ Failure in Trauma Patients <i>Yuqing Wang, Yun Zhao, Rachael Callcut, Linda Petzold</i>	25
Cloud Supportive Multi-Stage Selection Model for Multiple Imputation in Large Scale Study: Non-Ignorable NMAR <i>Jagan Mohan, Reddy Seelam, Srinivasa Reddy, G Vijaya Suresh</i>	41
Assessing long term degradation of industrial assets based on their production output curves <i>Pierre Dagnely, Peter Van Hese, Tom Tourwè, Elena Tsiporkova</i>	53
Detecting Anomalous Invoice Line Items in the Legal Case Lifecycle <i>Valentino Constantinou, Mori Kabiri</i>	69
Rymon Trees and Binary Data Biclusters <i>Joshua Yee, Haoyu Wang, Dan A. Simovici</i>	85
Two Strategies for the Classification of the Quality of Red Wine <i>Brian Benson, Warwick Graco, Hari Koesmarno, Ed Lewis, Garry Mitchell, Tony Nolan, Peter Phillips</i>	97
<i>Author Index</i>	115

21th Industrial Conference on Data Mining ICDM 2021

www.data-mining-forum.de

Chair

Petra Perner FutureLab Artificial Intelligence IBaI-2,
..... Germany

Program Committee

Ajith Abraham	Machine Intelligence Research Labs (MIR Labs), USA
Mohamed, Bourguessa	Universite du Quebec a Montreal - UQAM, Canada
Bernard Chen	University of Central Arkansas, USA
Antonio Dourado	University of Coimbra, Portugal
Jeroen de Bruin	Medical University of Vienna, Austria
Stefano Ferilli	University of Bari, Italy
Geert Gins	Glaxo Smith Kline, Belgium
Warwick Graco	ATO, Australia
Aleksandra Gruca	Silesian University of Technology, Poland
Pedro Isaias	Universidade Aberta (Portuguese Open University), Portugal
Piotr Jedrzejowicz	Gdynia Maritime University, Poland
Martti Juhola	University of Tampere, Finland
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Lui Xiaobing	Google Inc., USA
Mehmed Kantardzic	University of Louisville, USA
Eduardo F. Morales	INAOE, Ciencias Computacionales, Mexico
Samuel Noriega	Universitat de Barcelona, Spain
Juliane Perner	Cancer Research, Cambridge Institutes, UK
Moti Schneider	Netanya Academic College, Israel
Rainer Schmidt	University of Rostock, Germany
Victor Sheng	University of Central Arkansas, USA
Kaoru Shimada	Section of Medical Statistics, Fukuoka Dental College, Japan
Gero Szepannek	University Stralsund, Germany

Joao Miguel Costa Sousa
Markus Vattulainen
Zhu Bing

Technical University of Lisbon, Portugal
Tampere University, Finland
Sichuan University, China

Table of Content

BERTSurv: BERT-Based Survival Models for Predicting Outcomes of Trauma Patients <i>Yun Zhao, Qinghang Hong, Xinlu Zhang, Yu Deng, Yuqing Wang, Linda Petzold</i>	1
Method of Computing Similarity with Clusters <i>Arthur Yosef, Moti Schneider</i>	17
Empirical Analysis of Machine Learning Configurations for Prediction of Multiple Organ Failure in Trauma Patients <i>Yuqing Wang, Yun Zhao, Rachael Callcut, Linda Petzold</i>	25
Cloud Supportive Multi-Stage Selection Model for Multiple Imputation in Large Scale Study: Non-Ignorable NMAR <i>Jagan Mohan, Reddy Seelam, Srinivasa Reddy, G Vijaya Suresh</i>	41
Assessing long term degradation of industrial assets based on their production output curves <i>Pierre Dagnely, Peter Van Hese, Tom Tourwè, Elena Tsiporkova</i>	53
Detecting Anomalous Invoice Line Items in the Legal Case Lifecycle <i>Valentino Constantinou, Mori Kabiri</i>	69
Rymon Trees and Binary Data Biclusters <i>Joshua Yee, Haoyu Wang, Dan A. Simovici</i>	85
Two Strategies for the Classification of the Quality of Red Wine <i>Brian Benson, Warwick Graco, Hari Koesmarno, Ed Lewis, Garry Mitchell, Tony Nolan, Peter Phillips</i>	97