DATA ANALYTICS 2012

The First International Conference on Data Analytics

September 23-28, 2012

Barcelona, Spain

DATA ANALYTICS 2012 Editors

Sandjai Bhulai, VU University Amsterdam, The Netherlands

Joseph Zernik, Human Rights Alert (NGO), USA

Petre Dini, Concordia University, Canada / China Space Agency Center, China

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© (2012) by International Academy, Research, and Industry Association (IARIA)
Please refer to the Copyright Information page.

Printed by Curran Associates, Inc. (2012)

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: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

DATA ANALYTICS 1: TARGET ANALYTICS

Complexity Analysis in the Language of Information Technology Companies 1
Mary Luz Mouronte López

How to Find Important Users in a Web Community? Mining Similarity Graphs 10
Clemens Schefels

Characterization of Network Traffic Data: A Data Preprocessing and Data Mining Application 18
Esra Kahya-Özyirmidokuz, Ali Gezer, Cebrail Ciflikli

Structured Data and Source Code Analysis for Financial Fraud Investigations 24
Joe Sremack

DATA ANALYTICS 2: APPLICATION-ORIENTED ANALYTICS I

Integrity, or Lack Thereof, of the Electronic Record Systems of the Courts of the State of Israel 31
Joseph Zernik

Network Visualization of Car Inspection Data using Graph Layout 39
Jaakko Talonen, Miki Sirola, Mika Sulkava

Trend Visualization on Twitter: What's Hot and What's Not? 43
Sandjai Bhulai, Peter Kampstra, Lidewij Kooiman, Ger Koole, Marijn Deurloo, Bert Kok

Travel Time Estimation Results with Supervised Non-parametric Machine Learning Algorithms 49
Ivana Cavar, Zvonko Kavran, Ruder Michael Rapajic

Automated Predictive Assessment from Unstructured Student Writing 57
Norma C. Ming, Vivienne Ming

DATA ANALYTICS 3: APPLICATION-ORIENTED ANALYTICS II

Building OLAP Data Analytics by Storing Path-Enumeration Keys into Sorted Sets of Key-Value Store Databases 61
Luis Loyola, Fernando Wong, Daniel Pereira

Ontology-Guided Data Acquisition and Analysis 71
Dominic Girardi, Klaus Arthofer, Michael Giretzlehner

Analysis of Streaming Service Quality Using Data Analytics of Network Parameters 76
Jie Zhang, Hwa-Jong Kim, Doo-Heon Ahn

Modeling Team-Compatibility Factors Using a Semi-Markov Decision Process: A Data-Driven Framework for Performance Analysis in Soccer 80
Ali Jarvandi, Shahram Sarkani, Thomas Mazzuchi

Design and Operation of the Electronic Record Systems of the US Courts are Linked to Failing Banking Regulation 83
Joseph Zernik

DATA ANALYTICS 4: FUNDAMENTALS FOR DATA ANALYTICS

The Open Data Interface (ODI) Framework for Public Utilization of Big Data 94
Hwa-Jong Kim, Seung-Teak Lee, Yi-Chul Kang

Evaluating Data Minability Through Compression - An Experimental Study 97
Dan Simovici, Dan Pletea, Saaid Baraty

A New Measure of Rule Importance Using Hellinger Divergence 103
Chang-Hwan Lee

An Architecture for Semantically Enriched Data Stream Mining 107
Andreas Textor, Fabian Meyer, Marcus Thoss, Jan Schaefer, Reinhold Kroeger, Michael Frey

Extraction and Visualisation of Semantic Concepts from Document-Sets using Taxonomies (POSTER) ..... N/A
Andreas Schmidt, Daniel Kimmig, Robert Senger

Author Index