

Data Science & Social Research 2019

Book of Abstracts

Second international conference on data science and social
research

4 February 2019 - University of Milano - Bicocca

5 February 2019 - Università IULM

Editor: Paolo Mariani

Con il Patrocinio di



Regione
Lombardia



1^a edizione aprile 2019
Tutti i diritti riservati

© Copyright PKE srl
www.pkegroup.it
eMail: info@pke.it

ISBN: 9788894312096

Contents

Preface	11
Committees	12
Abstracts	13
Investigating vaccine sentiment in Italy over a period of ambiguous immunization policy	15
<i>Daniilo Ajovalasit, Veronica Dorgali, Angelo Mazza, Alberto D'onofrio and Piero Manfredi</i>	
Transitivity thresholds for Salo-Hamalainen index when the number of alternatives is greater than three	16
<i>Pietro Amenta, Antonio Lucadamo and Gabriella Marcarelli</i>	
Investigating the judges' performance in a national competition of sport dance . .	17
<i>Laura Anderlucci, Alessandro Lubisco and Stefania Mignani</i>	
Cognitive, social and psychological predictors of the graduate's disposition to entrepreneurship	18
<i>Pasquale Anselmi, Daiana Colledani, Luigi Fabbris and Egidio Robusto</i>	
Science mapping via dynamic topic modelling: an analysis on 30 years of social indicators research	19
<i>Massimo Aria, Michelangelo Misuraca and Maria Spano</i>	
Attractiveness of university degree programs: a social network analysis	20
<i>Silvia Bacci, Bruno Bertaccini and Alessandra Petrucci</i>	
Walkability assessment of urban areas through social media data mining	21
<i>Stefania Bandini, Andrea Gorrini and Giuseppe Vizzari</i>	
Extraction of cancer information from pathology clinical records using text mining	22
<i>Pietro Belloni, Giovanna Boccuzzo, Stefano Guzzinati, Irene Italiano, Bruno Scarpa, Carlo Rossi, Massimo Rugge and Manuel Zorzi</i>	
Spatial distribution of Multidimensional Educational Poverty using SAE	23
<i>Gaia Bertarelli, Caterina Giusti and Monica Pratesi</i>	
Social media data for social indicators: assessing the quality through case studies	24
<i>Silvia Biffignandi, Annamaria Bianchi and Camilla Salvatore</i>	
Google trends and Twitter: predictors or reactors? An application to bitcoin market determinants	26
<i>Silvia Biffignandi and Elena Pisanelli</i>	

The Role of Parties and Media for Income Inequality Perceptions. Evidence from the United States and the United Kingdom	27
<i>Chiara Binelli and Paul Matthew Loveless</i>	
Riskiness of Italian firms in the post-crisis period: an outlook through financial ratios	28
<i>Matilde Bini, Lucio Masserini and Alessandro Zeli</i>	
Media Bias and Crime Perception	29
<i>Riccardo Borgoni, Daniele Gualtieri and Alessandra Michelangeli</i>	
The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach	30
<i>Paolo Brunori and Guido Neidhofer</i>	
Mapping brand publics' social imaginaries on Instagram: how to use big data for exploring consumer culture	31
<i>Alessandro Caliendo and Guido Anselmi</i>	
Analysis of two-way ordinal contingency tables for social research	33
<i>Ida Camminatiello, Antonello D'Ambra and Luigi D'Ambra</i>	
Global optimization of a machine learning based forecasting pipeline	34
<i>Antonio Candelieri, Riccardo Perego and Francesco Archetti</i>	
Approximate dynamic programming for pumps scheduling optimization in urban water distribution systems	35
<i>Antonio Candelieri, Riccardo Perego and Iliaria Giordani</i>	
Modelling Housing Market Cycles in Global Cities	36
<i>Alessandra Canepa, Emilio Zanetti Chini and Huthaifa Alqaralleh</i>	
Divorce in Italy: a textual analysis of cassation judgments	37
<i>Rosanna Cataldo, Maria Gabriella Grassia, Marina Marino, Rocco Mazza and Vincenzo Pastena</i>	
Different sources of data for the Sustainability	38
<i>Rosanna Cataldo, Maria Gabriella Grassia, Marina Marino and Viktoriia Voitsekhovska</i>	
Mapping trending topics in social research methods	39
<i>Maria Carmela Catone, Paolo Diana, Giuseppe Giordano and Pierluigi Vitale</i>	
Composite indicators of the Soccer Players' Performance Indices	40
<i>Enrico Ciavolino, Maurizio Carpita and Paola Pasca</i>	
Modelling and testing on multivariate longitudinal data for nested design with application to player-by-player basketball analytics	41
<i>Livio Corain and Luigi Salmaso</i>	
Detection of mutimedia sexist contents	42
<i>Silvia Corchs, Elisabetta Fersini and Francesca Gasparini</i>	
Bayesian networks to discover similarities among subsets. Comparing European countries	43
<i>Rosario D'Agata, Simona Gozzo and Anna Maglia</i>	

Quality aspects when using mobile phone data in official statistics	44
<i>Fabrizio De Fausti, Roberta Radini, Luca Valentino and Tiziana Tuoto</i>	
A tourist tour planning for a smart tourism system in Calabria	45
<i>Annarita De Maio, Francesco Santoro and Antonio Violi</i>	
Web-based data collection and quality issues in co-authorship network analysis	46
<i>Domenico De Stefano, Vittorio Fuccella, Maria Prosperina Vitale and Susanna Zaccarin</i>	
Supervised and Unsupervised Hate Speech Detection	48
<i>Emiliano del Gobbo, Alice Tontodimamma, Lara Fontanella and Luigi Ippoliti</i>	
The Balance of Inequality: A Rediscovery of the Gini's R Concentration Ratio and a New Inequality Decomposition by Population Subgroups Based on a Physical Rationale	50
<i>Giorgio Di Maio and Paolo Landoni</i>	
Dimensionality reduction techniques on the simplex for text mining	51
<i>Simone Di Zio, Lara Fontanella, Sara Fontanella and Luigi Ippoliti</i>	
Ubu Roi. The Rise of Italian Populism through the Analysis of Social Media Content	53
<i>Ignazio Drudi, Fabrizio Alboni and Giorgio Tassinari</i>	
360° university governance through big data	54
<i>Angela Maria D'Uggento, Rosa Ceglie and Massimo Iaquinta</i>	
Comparing goal-based and result-based approach in modelling football outcomes	55
<i>Leonardo Egidi and Nicola Torelli</i>	
Pooling viewpoints to obtain a single evaluation. The ROI-MOB indicator of Erasmus+ mobility effects	57
<i>Luigi Fabbri and Manuela Scioni</i>	
A personalized smart tourism recommender system based on social media data	58
<i>Daniele Ferone, Elisabetta Fersini and Enza Messina</i>	
Excitable tweets: social computing and online sexism	59
<i>Antonia Anna Ferrante and Stamatia Portanova</i>	
Participants' behaviour at special events: sampling procedures and GPS technologies	61
<i>Mauro Ferrante, Domingo Martin-Martin and Stefano De Cantis</i>	
Automatic misogyny identification in online social media	62
<i>Elisabetta Fersini and Paolo Rosso</i>	
Digital news press mining through topic modeling, entity recognition and social networks analysis techniques	63
<i>Carlos G. Figuerola and Modesto Escobar</i>	
Spatial localization of mobile phone users and tourism flows in Sardinian destinations' network	64
<i>Annamaria Fiori and Ilaria Foroni</i>	

The challenges and limits of health open data in Italy	65
<i>Carlotta Galeone and Paolo Mariani</i>	
Grassroots-vs- influencers classification in analysis of Twitter news diffusion . . .	66
<i>Svitlana Galeshchuk and Ju Qui</i>	
Block-chain oriented system for the management of processes	68
<i>Massimiliano Giacalone, Emilio Massa, Diego Carmine Sinitò and Vito Santarcangelo</i>	
Practices and journeys: insights into environmental issues	69
<i>Paolo Giardullo</i>	
Machine learning approaches for prescription patterns analytics	70
<i>Ilaria Giordani, Gaia Arosio, Paolo Mariani, Ilaria Battiston, Antonio Candelieri and Francesco Archetti</i>	
Modelling human preferences by Bayesian optimization	71
<i>Ilaria Giordani, Attilio Redivo, Antonio Candelieri, Bruno Galuzzi and Francesco Archetti</i>	
On the improvement of soccer match result predictions	72
<i>Silvia Golia and Maurizio Carpita</i>	
GLAM organizations' Digital Maturity Indicator: a statistical approach for Campania Museums	73
<i>Massimo Guarino, Maria Anna Di Palma, Antonino Mario Olivieri and Michele Gallo</i>	
Passing Networks and Game Style in Football Teams: Evidences from European Champions League	74
<i>Riccardo Ievoli, Lucio Palazzo and Giancarlo Ragozini</i>	
A new and unified theory for time series models with ARMA representations and varying coefficients: one solution fits all	75
<i>Menelaos Karanasos, Alexandros Paraskevopoulos and Alessandra Canepa</i>	
Modeling and simulating durations of a professional tennis match	76
<i>Francesco Lisi and Matteo Grigoletto</i>	
The global health networks: a comparative analysis of tuberculosis, malaria and pneumonia using social media data	77
<i>Milena Lopreite, Michelangelo Puliga and Massimo Riccaboni</i>	
Basketball spatial performance indicators	79
<i>Marica Manisera, Rodolfo Metulini, Marco Sandri and Paola Zuccolotto</i>	
Predicting Cycling Usage for Improving Bike-Sharing Systems	80
<i>Giancarlo Manzi, Silvia Salini and Cristiano Villa</i>	
Gaming Analytics through Players (GAP). Profiling Italian Players	81
<i>Ilaria Mariani, Alan Mattiassi and Emma Zavarrone</i>	
Company requirements and monetary evaluation in the Italian healthcare industry	83
<i>Paolo Mariani, Andrea Marletta, Lucio Masserini and Mariangela Zenga</i>	
The risk of inappropriateness in the Italian geriatric wards using national Hospital Discharge Data	84
<i>Paolo Mariani, Andrea Marletta, Marcella Mazzoleni and Mariangela Zenga</i>	

A missing value approach on Facebook Big Data: Like, Dislike or Nothing?	85
<i>Andrea Marletta, Paolo Mariani and Erika Grammatica</i>	
A Global Rank of the Delphi survey items on the future scenarios of the Family.	86
<i>Marco Marozzi and Mario Bolzan</i>	
Social Media Disasters. Big data issues in public communication field	87
<i>Francesco Marrazzo and Gabriella Punziano</i>	
Human activity spatio-temporal indicators using mobile phone data	89
<i>Rodolfo Metulini and Maurizio Carpita</i>	
The influence of cities on intergenerational social mobility	90
<i>Alessandra Michelangeli and Umut Turk</i>	
A supervised learning approach in the risk estimate of gambling in adolescents: a case study	91
<i>Gianna Monti, Laura Benedan and Manuel Mercandelli</i>	
Regularized semiparametric estimation of high dimensional dynamic conditional covariance matrices	92
<i>Claudio Morana</i>	
Smart work as an employee welfare practice: an explorative research project	94
<i>Ariela Mortara</i>	
Virtual encounter-simulations: A new methodology for generating conflict data	95
<i>Georg Mueller</i>	
From recognition to re-use: open data for confiscated goods	97
<i>Giuseppe Notarstefano, Umberto Di Maggio and Giuseppe Ragusa</i>	
Promoting cruise ship as “tourist destination” on Television: the case of Italy	98
<i>Antonino Mario Oliveri and Gabriella Polizzi</i>	
Measuring tourist satisfaction and dissatisfaction: adaptation of the 4Q method- ology to the case of web based data.	99
<i>Antonino Mario Oliveri, Gabriella Polizzi, Anna Maria Parroco and Michele Gallo</i>	
Detecting multidimensional clustering across EU regions	100
<i>Pasquale Pavone, Margherita Russo, Francesco Pagliacci, Simone Righi and Anna Giorgi</i>	
Estimating high dimensional stochastic volatility models	101
<i>Matteo Pelagatti and Giacomo Sbrana</i>	
Exploring paths through placetellers performativity on Instagram	102
<i>Ilaria Primerano, Giuseppe Giordano and Pierluigi Vitale</i>	
Digital natives but not yet digital citizens: how the digital gap affects the educa- tional poverty of young people	103
<i>Luciana Quattrococchi and Gabriella Grassia</i>	
“Working smart” in a digital context: from digital technology to digital compe- tences	104
<i>Aurelio Ravarini</i>	

Analysing patterns and dynamics of the international student mobility networks in higher education. A comparative empirical study on European countries.	105
<i>Marialuise Restaino, Maria Prosperina Vitale and Ilaria Primo</i>	
A daily sentiment index on immigration of Italian-speaking Twitter users	107
<i>Alessandra Righi, Domenico Bianco and Mauro Gentile</i>	
Emotionographies of the city of Milan	108
<i>Elisabetta Risi</i>	
Human resources in neuromanagement: an added value to job assessment	110
<i>Vincenzo Russo, Margherita Zito and Giorgio Gabrielli</i>	
“The theatre of the mind”: the effect of radio exposition on tv advertising	112
<i>Vincenzo Russo, Alessandro Fici, Margherita Zito and Giorgio Gabrielli</i>	
A comparison of machine learning techniques for university students dropout prediction	114
<i>Annalina Sarra, Eugenia Nissi and Lara Fontanella</i>	
Personalized treatment of chronic diseases: models, experiments and preliminary results	115
<i>Fabio Sartori, Riccardo Melen, Matteo Lombardi and Davide Maggiotto</i>	
Social media big data: state of the art of some methodological challenges	117
<i>Andrea Sciandra</i>	
Classification supervised models to disclose action and information in “U.N. 2030 agenda” social media data	118
<i>Andrea Sciandra, Livio Finos and Alessio Surian</i>	
Protect your business: a model of profiling small businesses and freelancers on ability to secure online	119
<i>Bruno Giuseppe Sfogliarini</i>	
Matching on Poset based Average Rank for Multiple Treatments (MARMoT)	120
<i>Margherita Silan, Giovanna Boccuzzo and Bruno Arpino</i>	
Blending small and big data for evaluating the immigration-insecurity relationship	121
<i>Sonia Stefanizzi and Giancarlo Manzi</i>	
The Reuse of Digital Computer Data: Transformation, Recombination and Generation of Data Mixes in Big Data Science	123
<i>Niccolò Tempini</i>	
A preference index design for big data	124
<i>Venera Tomaselli and Giulio Giacomo Cantone</i>	
Measuring immigrants integration: simultaneous clustering and dimensionality reduction in PLS-SEM models	125
<i>Venera Tomaselli, Mario Fordellone and Maurizio Vichi</i>	
Destination branding and Residents: an analysis of the Valtellina territory	126
<i>Giovanni Tonini, Mariangela Zenga and Nadia Cominetti</i>	

Geocaching – Boxing Geosocial Traces	127
<i>Judit Varga</i>	
Part of Speech Tagging for blog and microblogs data	128
<i>Paola Zola and Silvia Golia</i>	
The “Sentability”: sentiment analysis for social media’s accountability. The case of environmental issues in Italian municipalities.	129
<i>Paola Zola, Laura Rocca, Davide Giacomini and Diego Paredi</i>	
Sponsors and Partners	131

Preface

As digital technologies, the internet and social media become increasingly integrated into society, a proliferation of digital footprints of human and societal behaviours are generated in our daily lives. All these data provide opportunities to study complex social systems, by the empirical observation of patterns in large-scale data, quantitative modelling and experiments. The social data revolution enables not only new business models but it also provides policy makers with better instruments to support their decisions. This conference aims at stimulating the debate between scholars of different disciplines about the so called “data revolution” in social research. Statisticians, computer scientists and domain experts in social research will discuss the opportunities and challenges of the social data revolution to create a fertile ground for addressing new research problems.

This book includes the abstracts of the papers presented at the second international conference on data science and social research whose authors paid the registration fee.

All abstracts appear in the book as received. Authors are responsible for the entire content and accuracy of their abstracts.

Committees

Scientific committee

Enrica Amaturò, Luigi Fabbris, Carlo Natale Lauro, Paolo Mariani, Enza Messina, Monica Pratesi, Sonia Stefanizzi, Nicola Torelli, Maurizio Vichi.

Programme committee

Biagio Aragona, Francesco Archetti, Carlo Batini, Chiara Binelli, Giovanna Boccuzzo, Federico Cabitza, Furio Camillo, Fabio Crescenzi, Corrado Crocetta, Elisabetta Fersini, Maria Gabriella Grassia, Filomena Maggino, Paolo Mariani, Marina Marino, Andrea Maurino, Enza Messina, Claudio Morana, Matteo Palmonari, Matteo Pelagatti, Sonia Stefanizzi, Rosanna Verde, Giuseppe Vizari, Biancamaria Zavanella, Emma Zavarrone.

Local organizing committee

Laura Benedan, Antonio Candelieri, Federica Codignola, Alessandra Decataldo, Elisabetta Fersini, Ilaria Giordani, Caterina Liberati, Paolo Mariani, Andrea Marletta, Marcella Mazzoleni, Grazia Murtarelli, Mauro Mussini, Debora Nozza, Matteo Pelagatti, Anisa Rula, Domingo Scisci, Sonia Stefanizzi, Emma Zavarrone, Mariangela Zenga.

A PREFERENCE INDEX DESIGN FOR BIG DATA

Tomaselli Venera¹ and Cantone Giulio Giacomo¹

¹ Department of Political and Social Sciences, University of Catania,
(e-mail: tomavene@unict.it, prgcan@gmail.com)

TripAdvisor is a business service that works as reputation system to guarantee quality in tourism experience. This kind of new services is based on Big Data technologies and characterized by generating, managing and summarizing, even with rating indexes, a quantitative experimental size of information, representing a frontier issue for data analysis.

These data are organized and offered to users by a filter system aimed to recommend consumer's choices. Through a methodological design oriented to reward competitive quality, this acts as a crowd-sourced evaluation system.

The stakeholders don't take the crowd-sourced evaluation passively as they still have a wide range of, reactive actions even with huge amounts of received reviews. They can re-organize their economic game of exposure to crowd-sourced reputation by adopting merely strategic and unsubstantial behaviours, or even frauds.

As past reviews and rating indexes provided by the websites can affect the building data process, in this paper we suppose that final information can be biased, leading itself into non-linear, asymmetric dynamics.

On the basis of an empirical study for approximately 26.000 scores on TripAdvisor multipoint scale organized into 8-years time series and harvested by *R* software, we propose a methodological design of a rating index. The index is robust for avoiding any manipulation of open-to-view-results multipoint scales and, at the same time, reflects the original aim to provide both a form of guarantee from risks of bad experience in tourism and a coherent ranking for benchmarking purposes.

KEYWORDS: crowd-sourced evaluation, TripAdvisor, multipoint scale, ranking system robustness.

References

- MASUM, H., & TOVEY, M. 2011. *The Reputation Society: How Online Opinions Are Reshaping the Offline World*, Cambridge: MIT Press.
- SAURO, J., & LEWIS, R. 2016. *Quantifying the User Experience. Practical Statistics for User Research, 2nd Edition*, Burlington: Morgan Kaufman, 249-276.

Sponsors

University of Milano - Bicocca
Department of Economics, Management and Statistics
Department of Sociology and Social Research
Department of Informatics, Systems and Communication

Partners

Applied Statistics Association
Associazione Italiana di Sociologia
Associazione Italiana di Ricerca Operativa
Bicocca Applied Statistics Center
Consorzio Milano Ricerche
Italian Statistical Society
PKE
Unidata Bicocca

Data Science & Social Research 2019

Book of Abstracts

Editor: Paolo Mariani

The book includes the abstracts accepted for presentation at the Data Science & Social Research 2019 conference, February 4-5, in Milan. The conference aims at stimulating the debate between scholars of different disciplines about the so called “data revolution” in social research. Statisticians, computer scientists and domain experts in social research will discuss the opportunities and challenges of the social data revolution to create a fertile ground for addressing new research problems.

ISBN: 9788894312096