

Big data analytics for smart cities

*Original*

Big data analytics for smart cities / Cerquitelli, T., Migliorini, S., Chiusano, S.. - In: ELECTRONICS. - ISSN 2079-9292. - ELETTRONICO. - 10:12(2021), p. 1439. [10.3390/electronics10121439]

*Availability:*

This version is available at: 11583/2930014 since: 2021-10-09T11:50:26Z

*Publisher:*

MDPI

*Published*

DOI:10.3390/electronics10121439

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# Big Data Analytics for Smart Cities

Tania Cerquitelli <sup>1,\*</sup>, Sara Migliorini <sup>2</sup> and Silvia Chiusano <sup>3</sup>

<sup>1</sup> Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy

<sup>2</sup> Department of Computer Science, University of Verona, 37134 Verona, Italy; sara.migliorini@univr.it

<sup>3</sup> Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, 10129 Torino, Italy; silvia.chiusano@polito.it

\* Correspondence: tania.cerquitelli@polito.it; Tel.: +39-011-090-7178

In the last few years, cities have become engines of wealth creation thanks to the advent of the new information and communication. The capability to both generate and collect the data of public interest within the urban area (e.g., information about social events, public service usage, and mobility) has increased at an unprecedented rate, to such an extent that data rapidly scales towards big (urban) data. Such abundance creates an unprecedented opportunity to understand the way people interact in and with the urban environment, and enables researchers to tackle important and urgent urban challenges (e.g., traffic congestion, air pollution, and energy sustainability) by adding intelligence to the urban environment.

The design and development of innovative services and solutions tailored to smart cities entails the acquisition, integration, and analysis of heterogeneous data (e.g., social network data, urban safety and security perception, mobility data, energy consumption data, and data that may increase citizen awareness of the urban environment). To collect, store, manage, and analyze data, as well as visualize the results of the data analysis process, in order to make them readable and usable by citizens, ubiquitous sensing technologies, advanced data management and analytics models, and novel visualization methods should be devised.

The goal for this Special Issue is to explore how newly emerging data analytics solutions and data mining algorithms can help the smart city to become smarter to keep citizens in touch with service providers for their involvement in city management and innovation.

This special issue includes seven papers selected after a thorough reviewing process. All papers, co-authored by 36 researchers (11 female and 25 male), addressed key research challenges in the context of smart cities. An overview of paper topics and proposed methodologies to address the selected research issues are shown in Figure 1. It includes a wordcloud generated through a state-of-the-art text-mining pipeline applied on titles and abstracts of all accepted papers. The top 50 frequent words are displayed. They highlight the macro-topics addressed (e.g., data, algorithm, services) in this special issue by offering a detailed characteristic of all data management and data analytics techniques (e.g., training, parameter, error, analytics) and the urban applications where they have been adopted (e.g., city, energy, hotel, quality, review).

The accepted papers can be categorized into two subgroups: (1) *Services to citizens*, including city similarities based on hotel reviews and ratings [1], studying the main correlations among external quality-influencing variables and actual metrics determining the quality of the espresso coffee [2], and free-floating car-sharing services driven by predictive algorithms [3], and (2) *Application-driven services*, including energy mapping [4], management of air quality [5], indoor localizations [6], and robots to develop sense-making [7]. A summary of each paper is provided below.



**Citation:** Cerquitelli, T.; Migliorini, S.; Chiusano, S. Big Data Analytics for Smart Cities. *Electronics* **2021**, *10*, 1439. <https://doi.org/10.3390/electronics10121439>

Received: 7 May 2021

Accepted: 1 June 2021

Published: 15 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

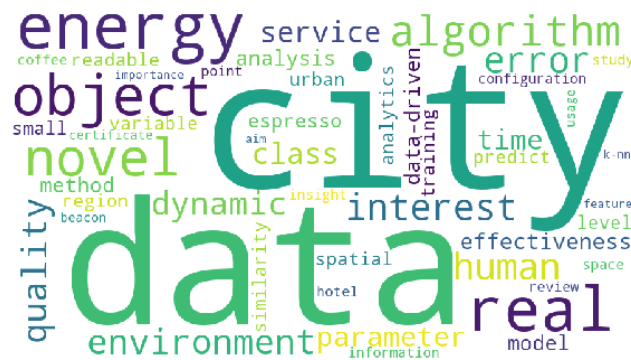


Figure 1. Special issue content: an overview.

**Services to citizens.** The authors of [1] address the challenge of investigating the similarities and dissimilarities between cities by considering the hotel reviews and ratings and their correlation with the surrounding Points of Interest (POIs). The paper proposes the use of a unified deep natural language processing (NLP) model, which analyses sentences in reviews and uses public TripAdvisor hotel-review datasets to validate the approach experimentally. The obtained results confirm the validity of the approach and suggest further investigation in this direction, for instance, by applying a multi-language approach that allows the consideration of reviews in different languages.

While coffee is one of the most popular beverages globally, the discussion of its quality is a controversial question, as discussed by the authors in [2]. Indeed, human experts usually perform its evaluation through the support of electronic and chemical tools. The authors propose the use of an association rule mining approach: by starting from a real-world dataset of espresso brewing by professional coffee-making machines, all correlations among external quality-influencing variables and actual metrics determining the quality of the espresso were extracted. The performed analysis was able to identify a set of interesting patterns and highlights how even in sub-optimal variables, we can obtain a high-quality product thanks to some compensation factors. The rules extracted by the analysis could be of great importance for domain experts and international coffee brands to improve their products.

Free-Floating Car-Sharing (FFCS) services are a flexible alternative to car ownership implemented in many big cities around the world. However, to spread the diffusion and adoption of such a service, it is necessary to predict its demand patterns over time and space, as proposed by the authors of [3]. Through the use of a real FFCS dataset regarding the city of Vancouver enriched with some socio-demographic information, the paper tries to predict such usage patterns by comparing several machine learning algorithms in terms of accuracy and ease of training. The performed analysis shows that it is possible to predict the future usage with relative errors down to 10%, while the spatial prediction can be estimated with relative errors of about 40%. In both cases, the best models are obtained through a Random Forests Regression after identifying the right set of features. Their identification also represents an essential contribution of the paper since it allows one to understand the FFCS system better and provide a high-quality service for both providers and customers.

**Application-driven services.** The authors in [4] proposed TUCANA, a data-driven engine to derive dynamic high-resolution geospatial maps by exploiting energy performance certificates released as open data. They presented a self-tuning cluster analysis to group buildings with similar properties on geometric and energy efficiency features and date of construction. The outcomes of the analytics step are exploited to plot cluster-marker maps, an innovative data visualization technique able to represent in a 2D map the high-dimensional results provided by the multivariate analysis. Furthermore, the authors in [4] proposed an innovative spatial constrained K-NN algorithm to infer different energy-related features for buildings without energy certificates and/or predict missing

values/noisy data. TUCANA has been validated on open data released by the Piedmont Region and integrates an interactive web application to provide navigable maps tailored to different stakeholders.

The management of air quality is one of the most important concerns for smart cities as discussed in [5]. The increasing availability of urban data has motivated the use of big data analytic techniques for making predictions and improving the air quality of a city. However, the availability of massive datasets does not always imply having objective and complete truths about the current situation. It has been observed that the collected data are often big but biased. For such reasons, this paper proposes a technique for correcting the bias present in real datasets in order to making better predictions. Moreover, the problem of automatic bandwidth selection was addressed through a bootstrap algorithm. The proposed technique was applied to a real-world case study regarding the city of A Coruña (Galicia, NW Spain).

Within urban environments, robots can carry out activities to help humans in their daily activities (e.g., safety monitoring, pre-emptive elderly care, and door-to-door garbage collection). To this aim, the robots should easily adapt to the surrounding environments and be equipped with algorithms to develop sense-making. To this aim, as discussed in [7], a basic requirement is the capability of robots to be equipped with predictive algorithms able to be trained with a few shot images. A detailed experimental comparison of state-of-the-art methods for deep few-shot image matching is performed on an innovative task-agnostic data set, including 2D views of openly available data, ShapeNet, and Google Images, properly combined. The top performance is obtained through the Siamese CNN with L2 normalizing the embedding, which allows to the correctly matching of images within the same visual domain.

The continuous innovation process of smart cities keeps pace with the innovation of the Industry 4.0 environment, since both scenarios have different aspects in common, and contributions in the first scenario have a significant impact on the second. The authors in [6] studied indoor localization's problem through commercial, off-the-shelf, low-cost BLE transmitters and receivers, easily deployed and integrated into real-life settings. Although the study is performed in an industrial scenario, it can be easily adapted to any indoor environment. Among the state-of-the-art localization algorithms, experiments confirmed that fingerprinting-based methods reduce the positioning error compared to distance-based methods. However, the authors recommended selecting the K-NN algorithm due to its simplicity and flexibility.

Cities are engines of wealth creation, yet rapid urbanization comes with side effects. Cities are greatly affected by crime, diseases, and pollution, significantly deteriorating the quality of life of their dwellers. Data-driven methodologies are playing an important role to deeply understand relevant issues to be addressed. However, to add more intelligence in the urban environment, a few open issues, which can be seen both as challenges and as opportunities simultaneously for both researchers and practitioners, are summarized below.

*Human-readable predictive models* As predictive algorithms increasingly support different aspects of our life, especially in a smart city scenario, a greater level of transparency is badly needed, not least because discrimination and biases have to be avoided. Increased transparency in the data-driven methodology is the first step towards the vision of "shared governance" that can lead to urban environments and policies that are relevant to the city-dwellers.

*Towards open science and transparent cities.* Data-driven methodologies model both urban environments and citizens' movements and activities by analyzing data. Unfortunately, only a limited number of high-quality and open datasets are available. Open science is the first step towards transparent cities. Thus, service public administration, service providers, and government bodies must release the collected data within urban environments to allow researchers and practitioners to analyze them and integrate the obtained results to create circular and positive global outcomes.

We hope that readers will find the issue of interest, and that its content will inspire future research activities.

**Acknowledgments:** The guest editors thank all those who contributed to the success of the special issue. First of all, a special thank goes to all the authors attracted by the Special Issue topics who submitted their research contributions. Many thanks go to the reviewers who offered their expertise and valuable effort in promptly evaluating the submitted papers, providing comprehensive, critical, and constructive comments to the authors by ensuring high-quality research standards of the special issue. Moreover, a great appreciation goes to journal editors timely support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cagliero, L.; La Quatra, M.; Apiletti, D. From Hotel Reviews to City Similarities: A Unified Latent-Space Model. *Electronics* **2020**, *9*, 197. [[CrossRef](#)]
2. Apiletti, D.; Pastor, E. Correlating Espresso Quality with Coffee-Machine Parameters by Means of Association Rule Mining. *Electronics* **2020**, *9*, 100. [[CrossRef](#)]
3. Cocca, M.; Teixeira, D.; Vassio, L.; Mellia, M.; Almeida, J.M.; da Silva, A.P.C. On Car-Sharing Usage Prediction with Open Socio-Demographic Data. *Electronics* **2020**, *9*, 72. [[CrossRef](#)]
4. Cerquitelli, T.; Di Corso, E.; Proto, S.; Bethaz, P.; Mazzarelli, D.; Capozzoli, A.; Baralis, E.; Mellia, M.; Casagrande, S.; Tamburini, M. A Data-Driven Energy Platform: From Energy Performance Certificates to Human-Readable Knowledge through Dynamic High-Resolution Geospatial Maps. *Electronics* **2020**, *9*, 2132. [[CrossRef](#)]
5. Borrajo, L.; Cao, R. Big-But-Biased Data Analytics for Air Quality. *Electronics* **2020**, *9*, 1551. [[CrossRef](#)]
6. Cannizzaro, D.; Zafiri, M.; Jahier Pagliari, D.; Patti, E.; Macii, E.; Poncino, M.; Acquaviva, A. A Comparison Analysis of BLE-Based Algorithms for Localization in Industrial Environments. *Electronics* **2020**, *9*, 44. [[CrossRef](#)]
7. Chiatti, A.; Bardaro, G.; Bastianelli, E.; Tiddi, I.; Mitra, P.; Motta, E. Task-Agnostic Object Recognition for Mobile Robots through Few-Shot Image Matching. *Electronics* **2020**, *9*, 380. [[CrossRef](#)]