EDITORIAL

Introduction to the special issue on social mining and big data ecosystem for open, responsible data science

Luca Pappalardo[1] · **Valerio Grossi**[1] · **Dino Pedreschi**[2]

Data Science is rapidly changing the way we do business, socialize, and govern society. It is also changing the way scientific research is performed. A new paradigm is emerging, where theories and models and the bottom-up discovery of knowledge from data mutually support each other. Experiments and analyses over massive datasets are becoming functional not only to the validation of existing theories and models but also to the data-driven discovery of patterns emerging from data, which can help scientists design better theories and models, yielding a deeper understanding of the complexity of social, economic, biological, technological, cultural and natural phenomena. Data Science, intertwined with Artificial Intelligence, emerged as a disruptive consequence of the digital revolution. If appropriately oriented at social good and human values, Data Science and AI can help tackle the global challenges facing humanity, well represented in the SGDs, the Sustainable Development Goals set forth by the United Nations, and dramatically highlighted by the pandemic.

To fully reap the fruits of Data Science for social good, though, it is necessary to scale up the capacity of interdisciplinary researchers and innovators to take advantage of the new powerful tools and methods of Data Science and AI. Therefore, we need integrated ecosystems for ethic-sensitive scientific discoveries and advanced applications of social data mining applied to the various dimensions of social life, as recorded by "big data". To the purpose of boosting data-driven research in multiple fields, including human, social and economic sciences, by enabling easy comparison, re-use, and integration of state-of-the-art big social data, methods, and services, into new socially impactful research. Social Mining and Big Data Ecosystems for Open, Responsible Data Science are necessary, which not only strengthen the existing clusters of excellence in social data mining research but also strive at creating global, interdisciplinary communities of social data scientists, fostered by extensive training, networking, and innovation activities. Such ecosystems are based on three pillars: infrastructures granting access to large datasets, analytical and AI tools, and data-driven experiments; communities of data scientists and AI experts; broad communities of users and stakeholders.

One initiative that exemplifies the approach is SoBig-Data++ ,the Integrated Infrastructure for Social Mining & Big Data Analytics supported since 2015 by the European Commission under the H2020 program "Excellent Science—Research Infrastructures''. SoBigData++ is a research infrastructure for open data science, at the second stage of "Advanced community", aggregating hundreds of interdisciplinary scientists from 31 partners of 14 EU countries and open to international scientists and innovators (http://www.sobigdata.eu/). We are thankful to the SoBig-Data++ community for providing many excellent concrete examples of socially relevant and impactful data-driven research, also showing how the idea of a Social Mining and Big Data Ecosystems for Open, Responsible Data Science can boost a transformative effect in basically all scientific disciplines. Several of these examples are described in this special issue, which was launched with the idea of soliciting contributions from researchers and practitioners in data mining and other disciplines to share their research in big data analytics and data science applications. The special issue was designed to target contributions using data mining and machine learning on social data to tackle socially relevant challenges in original and ethical ways, from data collection to model exploitation. Another objective of the special

✉ Dino Pedreschi
dino.pedreschi@unipi.it

Luca Pappalardo
luca.pappalardo@isti.cnr.it

Valerio Grossi
valerio.grossi@isti.cnr.it

[1] Institute of Information Science and Technologies, National Research Council of Italy (ISTI-CNR), Pisa, Italy

[2] Department of Computer Science, University of Pisa, Pisa, Italy

issue was to understand if and how the emerging research infrastructures and the associated data science research eco-systems are enhancing the capacities of social data-driven research.

It is our great pleasure to introduce the resulting collection of papers in the special issue on "Social Mining and Big Data Ecosystem for Open, Responsible Data Science". We are excited to witness amazing progress and a growing global community of scientists, and we hope that this special issue succeeds in providing an honest and passionate account of a global scientific trend that is profoundly transforming science, providing it better means to foster social good.

We thank the Editor-in-Chief of the International Journal of Data Science and Analytics (JDSA), Professor Longbing Cao, for the opportunity to guest-edit this collection. We are also very thankful to the contributing authors and the reviewers who carefully examined the papers. We ended accepting six papers, covering a wide spectrum of challenging issues:

- *Data Science: a game-changer for science and innovation*: a paper showing how data science impacts science and society at large, including ethical and governance issues connected with managing data that touch upon aspects of human behaviour.
- *Measuring objective and subjective well-being: dimensions and data sources*: a paper that illustrates the approaches for measuring well-being. The authors distinguish between objective and subjective well-being and surveys the theoretical background, the relevant dimensions of well-being, the new data sources for measurement, and relevant recent studies.
- *(So) Big Data and the transformation of the city*: this paper discusses the main issues of urban data analytics, focusing on privacy issues, algorithms, applications, and georeferenced data from social media. The authors leverage, as concrete case studies of urban data science tools, the results obtained in the "City of Citizens" thematic area of the SoBigData initiative, which includes a virtual research environment with mobility datasets and urban analytics methods developed by several institutions around Europe.
- *Human migration: the big data perspective*: in this paper the authors answer the question "How can big data help to understand the migration phenomenon?" through an analysis of various phases of migration, comparing traditional and novel data sources and models at each phase. They focus on three phases of migration—the journey, the stay, and the return—at each phase describing the state of the art and recent developments and ideas.
- *A Workflow Language for Research e-Infrastructures*: this paper outlines the HyWare language and platform. The language is an extension of the traditional workflow languages enabling the definition of workflows including automatic and manual analytical steps with the purpose of replicating and building large-scale data-driven experiments.
- *An ethical and legal framework for social data science*: this paper provides a framework for research infrastructures enabling ethically-sensitive and legally compliant data science, helping data scientists to frame the appropriate self-assessment questions to ensure an ethical, responsible design, implementation and deployment of data science projects.

We, as guest editors, are proud to offer this collection to the attention and scrutiny of the scientific community.