



Editorial

Editorial for the Special Issue on “Software Engineering and Data Science”

Davide Tosi

Department of Theoretical and Applied Sciences (DiSTA), Università degli Studi dell'Insubria, 21100 Varese, Italy; davide.tosi@uninsubria.it

In the last few years, data-driven software solutions have attracted a lot of attention in research and development at academic, industry, business, and government levels to exploit the hidden knowledge and big data that can be offered to cities and citizens in the future. However, data-driven software solutions are different from “traditional” software development projects, as the focus of the main development core is on managing data (e.g., data store and data quality) and designing behavioral models with the aid of artificial intelligence and machine learning techniques. To this end, new life-cycles, algorithms, methods, processes, and tools are required. The Special Issue, “Software Engineering and Data Science”, in the Journal of Future Internet, is devoted to recent trends and advancements in the field of engineering data-intensive software solutions to address challenges in developing, testing, and maintaining such data-driven systems. We received 13 submissions; after the initial screening and the peer review process, six papers have been finally accepted for publication. Accepted articles can be classified into two sets: (1) application of data-driven solutions to real-life problems and (2) techniques and algorithms addressing the different challenges of data-driven software engineering.

The first set of articles discusses the applicability of data science and data-driven solutions to everyday problems. Casini et al. [1] studied the inversion in the decreased/increased rate of new SARS-COV-2 infections in the countries involved in the European football championship that took place from 11 June to 11 July 2021, investigating the hypothesis of an association. They collected and analyzed all data regarding COVID-19 infections from the official online repositories. Then, they adopted Bayesian piecewise regression with a Poisson generalized linear model to look for changepoints in the time series of the new SARS-COV-2 cases of each country involved in the 2020 European football championship. For all the 17 countries involved, the changepoint coincides with an inversion in the SARS-COV-2 case rate from a decreasing to an increasing rate of infections, thus suggesting an association between infection rates and the European football championship. Another example of applying data science to real-life is presented in the work of Tosi et al. [2]. They conducted a correlation study using heterogeneous data sources, such as Google mobility data, SARS-COV-2 infection data, and the official dataset relating to infections in Italian schools for the period of 14 September 2020–30 October 2020. Three extensive Italian regions (Lombardy, Campania, and Emilia) (that adopted different approaches in opening and closing schools to contrast infections) have been deeply studied to understand the main driver that sparked the second SARS-COV-2 wave in Italy. The conducted data analyses suggest that schools are a driver of contagion and are not a safe environment by definition. Munjal et al. [3] applied big data-driven solutions to smart cities. Smart cities will be equipped with millions of smart devices and network connections, thus requiring a high level of energy consumption and carbon emissions. The authors defined a public transport-assisted data-dissemination system to utilize public transport as another communication medium, along with other networks, with the help of software-defined technology. The main objective is to minimize energy consumption with maximum data delivery. To this end, a multi-attribute decision-making algorithm is designed to self-identify



Citation: Tosi, D. Editorial for the Special Issue on “Software Engineering and Data Science”. *Future Internet* **2022**, *14*, 306. <https://doi.org/10.3390/fi14110306>

Received: 24 October 2022

Accepted: 24 October 2022

Published: 26 October 2022

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



Copyright: © 2022 by the author. 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/>).

the best network among wired, wireless, and public transport networks based on users' requirements and different services. Once public transport was selected as the best network, the Capacitated Vehicle Routing Problem (CVRP) will be implemented to offload data onto buses as per the maximum capacity of buses.

The second set of articles discusses new development methodologies, algorithms for software libraries recommendation, and technologies for ontology-based knowledge extraction from various heterogeneous sources. Almeida et al. [4] addressed the combined adoption of Agile and DevOps software development methodologies to cope with the increasing complexity of managing customer requirements and development requests. The authors presented a qualitative methodology to analyze the benefits that can arise from the combination of the two methodologies. A comprehensive set of twelve case studies, representing practices of the simultaneous adoption of both methodologies, was assessed. The simultaneous adoption of Agile and DevOps, when properly combined and aligned, allows (1) developers to gain greater control over the environment, infrastructure, and applications; (2) a more collaborative and Agile framework; (3) to simplify and automate the model processes to make them more rational and efficient. Krasanakis et al. [5] studied how to help developers automatically discover libraries to be reused in their software projects. They extended the accurate project-library recommendation systems, which employ Graph Neural Networks, with a revised collaborative graph filtering mechanism. The revised filtering mechanism exploits partially absorbing random walk filters, which the authors theorized could emulate human-driven library discovery. The experimental results on a real-world dependency graph of Android project third-party library dependencies highlighted promising research directions in automated software engineering and broader collaborative filtering research. Sikelis et al. [6] provided insight into critical aspects of ontology-based knowledge extraction from various heterogeneous sources, such as text, databases, and human expertise, realized in feature selection. Ontology-based algorithms and approaches are described to represent features and perform feature selection and classification. Moreover, the authors highlighted open issues and challenges related to the research topic of ontology-based knowledge extraction.

We would like to thank all the authors for the papers they submitted to this Special Issue. We would also like to acknowledge all the reviewers for their careful and timely reviews which helped to improve the quality of this Special Issue.

Funding: This research received no external funding.

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

References

1. Casini, L.; Roccetti, M. A bayesian analysis of the inversion of the SARS-COV-2 case rate in the countries of the 2020 European Football Championship. *Future Internet* **2021**, *13*, 8. [\[CrossRef\]](#)
2. Tosi, D.; Campi, A.S. How schools affected the COVID-19 pandemic in Italy: Data analysis for Lombardy Region, Campania Region, and Emilia Region. *Future Internet* **2021**, *13*, 5. [\[CrossRef\]](#)
3. Munjal, R.; Liu, W.; Li, X.; Gutierrez, J.; Chong, P.H.J. Multi-Attribute decision making for energy-efficient public transport network selection in smart cities. *Future Internet* **2022**, *14*, 42. [\[CrossRef\]](#)
4. Almeida, F.; Simões, J.; Lopes, S. Exploring the benefits of combining DevOps and Agile. *Future Internet* **2022**, *14*, 63. [\[CrossRef\]](#)
5. Krasanakis, E.; Symeonidis, A. Fast library recommendation in software dependency graphs with symmetric partially absorbing random walks. *Future Internet* **2022**, *14*, 124. [\[CrossRef\]](#)
6. Sikelis, K.; Tsekouras, G.E.; Kotis, K. Ontology-based feature selection: A survey. *Future Internet* **2021**, *13*, 158. [\[CrossRef\]](#)