Peter Haber
Thomas J. Lampoltshammer
Helmut Leopold
Manfred Mayr *Eds.*

Data Science – Analytics and Applications

Proceedings of the 4th International Data Science Conference – iDSC2021

EBOOK INSIDE

Springer Vieweg

Peter Haber · Thomas J. Lampoltshammer ·
Helmut Leopold · Manfred Mayr
Editors

Data Science – Analytics and Applications

Proceedings of the 4th International Data Science Conference – iDSC2021

Springer Vieweg

Editors
Peter Haber
Salzburg University of Applied Sciences
Salzburg, Austria

Thomas J. Lampoltshammer
University for Continuing Education Krems
Krems, Austria

Helmut Leopold
AIT Austrian Institute of
Technology GMBH
Vienna, Austria

Manfred Mayr
Salzburg University of Applied Sciences
Salzburg, Austria

ISBN 978-3-658-36294-2 ISBN 978-3-658-36295-9 (eBook)
https://doi.org/10.1007/978-3-658-36295-9

Responsible Editor: David Imgrund
This Springer Vieweg imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH part of Springer Nature.
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Preface

Based on the overall digitalisation in all spheres of our lives, Data Science and Artificial Intelligence (AI) are nowadays cornerstones for innovation, problem solutions and business transformation. Data, whether structured or unstructured, numerical, textual, or audiovisual, put in context with other data or analysed and processed by smart algorithms, are the basis for intelligent concepts and effective solutions. These solutions are addressing many application areas such as Industry 4.0, Internet of Things (IoT), smart cities, smart energy generation and distribution, and environmental management. Innovation dynamics and business opportunities as effective solutions for the essential societal, environment, or health challenges, are enabled and driven by modern data science approaches.

However, Data Science and Artificial Intelligence are forming a new field that needs attention and focused research. Effective data science is only achieved in a broad and diverse discourse – when data science experts cooperate tightly with application domain experts and scientists exchange views and methods with engineers and business experts. Thus, the **4th International Data Science Conference** (iDSC 2021) brought together researchers, scientists, and business experts to discuss new approaches, methods, and tools made possible by data science.

The cooperation of the Salzburg University of Applied Sciences, the Vorarlberg University of Applied Sciences, the University for Continuing Education Krems (Danube University Krems), and the AIT Austrian Institute of Technology demonstrates the strong Austrian scientific footprint and a deep commitment to a cooperative effort for jointly building an international community from science, research, as well as business, through data science and data analytics.

The iDSC is designed as a conference with a dual approach: By bringing together the latest findings in data science research and innovative implementation examples in business and industry, the conference is aimed at reflecting the current scientific breakthroughs and application expertise, as a means of stimulating shared professional discourse. The six thematic sessions of the conference have been mirroring all the top issues of the data science discipline ranging from challenges in the industrial setting, via Deep and Machine learning methodologies and Natural Language Processing approaches up to future innovation strategies. While the Research Track had a strong emphasis on Safety and Security matters like Anomaly Detection, Integrity Awareness or Ethical fairness of AI e.g., the Industry Track more made deep reflections on the Scaling of Business models, Infrastructure and Software, the Use of Data Science in SMEs, or the development of Data Ecosystems as an Incubator for Data Innovation.

5 keynotes, two from the research arena and three from different industries, had enhanced the conference's total outcome by deliberating on data science from a meta-level: The topics "AI everywhere as a Social good", "The paradigm shift to the circular economy pushed by Smart Data", "The new cultural mindset of Data Mesh", "Open Science with limited closed data-sharing and the development of synthetic data models" and last but not least "Potentials for agile and transparent Data Governance with domain-driven, decentralized Data management" are representing a shining mix of themes – philosophically grounded – making the iDSC an event not to miss in the community.

Therefore - once again - the iDSC, showed that the chosen structure of research and industry tracks provided fascinating insights into current areas of research, as well as presenting impressive use cases to demonstrate the huge potential and significance of modern data science. With our new service, all talks can be now accessed also via video stream from our landing page https://idsc.at.

This was also made possible by the excellent work and outstanding execution of the conference by Michael Mürling, Athina Lykou and the team at AIT, as well as by our technical support team brutkasten.com. We greatly appreciate the generous support of SBA Research, X-NET Services, SPARX Services and as well our media partner Trending Topics. We also want to acknowledge Julian Nöbauer and Maximilian Tschuchnig from the Salzburg University of Applied Sciences for their ongoing backing especially on the creation of the proceedings. Finally, we thank the organizing & program committee as well as the reviewers for helping to make the conference a reoccurring success.

Enjoy the present proceedings of the conference and see you in 2023 at the University for Continuing Education Krems (Danube University Krems).

Peter Haber, Thomas Lampoltshammer, Helmut Leopold, Manfred Mayr
Conference Chairs

Data science & AI depend on smart ecosystems to provide society with innovative solutions

An overview of AI solutions "Made in Austria"

Digitalisation has changed the rules of business and many social mechanisms at an amazing pace. While hugely powerful devices such as smartphones, laptops, and PCs have served to network people on a global scale during the past decade, this transformation process has gained further momentum through the networking of our physical objects to create the Internet of Things (IoT). These developments, in turn, create the potential for new applications, business models, and value chains. However, this has simultaneously made us dependent on technology platforms, to the extent that our economy, our social lives, and our public administration are now all unthinkable without functioning digital infrastructures.

Three challenges must be overcome to ensure that this transformation is beneficial for mankind:

1. Mastery of digital technology platforms has become a fundamental requirement for business and society. Digital technology and infrastructure must be designed for maximum availability and offer the best possible level of security from a wide range of threats. Developments which focus on both minimal resource consumption and data protection in the service of humanity are essential.

2. Establishing extremely high-performance data management and ensuring that data sovereignty remains in the hands of the user is the order of the day. By living and working with a multitude of IoT devices, we continually generate huge data volumes which can be combined and processed using smart algorithms to produce essential information. This allows us to use digital platforms and smart data management to effectively address society's key challenges, including the environment, energy, and mobility. Smart IT services and high-performance computing (HPC) play a key role in determining productivity in our digital future. Contrary to the current cloud megatrend, this will also require new network architectures to balance data transmission and computing power between end devices (IoT) and data centres (cloud).

3. Lastly, effective cooperation must be fostered between data scientists and domain experts. Effective and solution-oriented data science and artificial intelligence (AI) can only function based on new forms of cooperation between the various disciplines. Computer scientists rely on mechanical engineers, electrical engineers, physicists, architects, etc., and vice versa, to successfully develop useful, needs-based, functional data science and AI solutions.

These challenges are a key research focus at the Center for Digital Safety & Security at the AIT Austrian Institute of Technology. Current highly innovative AIT developments "Made in Austria" include:

- **Smart encryption for secure cloud solutions:**

  » Smart data encryption to give data owners dedicated selective and dynamic access, even in distributed cloud systems (e.g., https://secredas-project.eu/ and https://profet.at/).

  » Next-generation data back-up and archiving solutions in public or hybrid cloud storage. Distributed and encrypted data can be stored securely in the cloud, without even the cloud provider being able to access and analyse the stored data (https://www.fragmentix.com/de/).

» Virtualized, distributed (blockchain-based) database architectures in the cloud to create new marketplaces. Using a highly secure cloud solution, encrypted supply and demand information in a distributed system can be retrieved automatically and compared anonymously (https://www.flexprod.at/de). This innovation won the German Digital Leader Award 2020 and is marketed through CATCH.direct (https://www.catch.direct/).

- **Cyber security AI solutions and new quality of experience for customers of digital services:**

  » Modern cyber security solutions must also be able to detect unknown and non-specified threats and attacks on IT systems. New AI-based anomaly detection systems are therefore essential for future security information and event management systems (SIEM) (https://aecid.ait.ac.at/), H. Leopold et al. Cyber Attack Information System – Erfahrungen und Erkenntnisse aus der IKT-Sicherheitsforschung, 2015, Springer Verlag, https://link.springer.com/book/10.1007/978-3-662-44306-4)

  » Tomorrow's software and system development need new software engineering approaches, particularly for safety-critical systems. This ensures safety & security are factored into the design, allowing effective security certification (https://www.threatget.com/).

  » Modern network operators need machine learning systems for effective and dynamic network management and to provide the best possible quality of experience to digital end users (https://bigdama.ait.ac.at/).

  » Effective AI solutions for protection against cyber crime (https://www.fakeshop.at/) and as weapons in the battle against disinformation and fake news (https://www.defalsif.ai).

- **Artificial intelligence and new data economies in the service of mankind as an important contribution to solving important societal challenges:**

  » Access to the data continually being generated in the digital space (Open Data) as well as new IT system architectures and algorithms are needed to enable new data economies and to support data cooperation and data sharing. Examples include the Austrian Data Intelligence Offensive (DIO) (https://www.dataintelligence.at/) and the European Gaia-X initiative.

  » The general availability of data about product characteristics, materials and life cycles allows raw materials to be recycled and products to be specifically processed in keeping with the concept of a circular economy.

  » Shifting AI from the data centre to the edge gives rise to numerous new applications while simultaneously increasing resilience and data security and reducing overall energy consumption.

All these examples show what fascinating innovations are possible when data experts cooperate closely with domain experts, users and authorities and share their expertise in innovation processes and smart, agile design-thinking ecosystems. Ultimately, smart ecosystems are the true drivers of innovation when it comes to developing AI solutions that will benefit humanity.

Helmut Leopold
Head of Center for Digital Safety & Security
AIT - Austrian Institute of Technology

Data boost industry-academia link

The bilateral focus of the iDSC conference is nicely reflected in the progress and results of DataKMU, which is a three-year research and transfer endeavour with participation from the industry as well as from academia including Salzburg University of Applied Sciences. The wide-spread uncertainty in small and medium-sized enterprises as regards utilisation of current data-driven concepts and methodologies to improve their businesses has motivated the DataKMU consortium to provide low-threshold access to a wide variety of state-of-the-art approaches in applied data science. The central goal is to systematically establish a multi-faceted operational industry-academia link in the field of data science, which is presented in the following.

A pivotal step of improvement in any situation is to get a clear picture of one own‘s status quo and to derive from this insight well-suited potential options for enhancement. Therefore, a multi-case study was undertaken to infer criteria for the determination of a so-called Data Science Readiness Level specifically for SMEs. This multidimensional scale, which is in part based on the Data Science Maturity Model from Oracle (described in the article “Strategic Approaches to the Use of Data Science in SMEs” in this Proceedings), allows SMEs to position themselves with little effort and to derive from this proper options for initial quick-wins in advanced data utilisation. This can be considered ramp-up support. SMEs can also locate themselves inside groups such as Practitioners, Strategists, and Pioneers and get related hints and information regarding suggested action points and potential caveats.

Data science strategies and their technical implementation are usually very domain-specific, which is why several showcase prototypes were created as part of the DataKMU project. These tangible results are open to analysis and discussion from various stakeholders, which is specifically interesting for companies that are too small to have their own data science teams on their payroll. The best-practise example implementations such as (i) Transfer Learning for automated data labelling in marketing, (ii) Data Analytics as a Service for virtual sensors as system observer in production, (iii) Development of automated methods for the detection and data extraction of signposts in tourism, and (iv) Automated recording of road conditions in logistics motivate SMEs to adopt similar solutions to stay competitive in their respective markets. Thus, by acting as ‚innovation followers‘ in the beginning, these SMEs can lower their threshold to the level of innovation leaders considerably.

In addition to the above-mentioned example implementations, a rather generic big data pipeline infrastructure was established to support SMEs in their bootstrapping of their own data science applications. This pipeline is a stateful pipeline that ‘remembers’ parametrisations over time in a way that can be re-played and adopted to optimise the various algorithmic building blocks. Thus, variants of solutions and related success can be analysed easily without the usual hardware costs.

The joint operation of several tertiary educational institutions helps to sharpen the respective profiles of data science-related programs in various places in the western part of Austria and adjacent regions of Germany. Replicating similar curricula is not an option to encourage a maximum of potential students to enrol in MINT programs. Thus, the specific strengths of the institutions were identified in course of the DataKMU project and also the main present and future research focus areas representing their characteristics in the individual universities. It is by far better to develop specific fields of application to address a higher total number of young people with study interest in data science. The intended heterogeneity in education also has a broader publicity as a side effect.

One of the pivotal results of the DataKMU endeavour is that the linkage between regional business players and regional universities is a very productive setting for the innovation system. Local businesses domain knowledge combined with methodological competences from research institutions enriched by the creative potential of young students forms an incubator setting that is beneficial to all the various stakeholders likewise. This creates a faster convergence of solutions in a pre-competitive environment.

In addition, a long-term strategic roadmap was developed together with the consortium by an external partner. For this purpose, a catalogue of measures was developed with external expertise to sustainably anchor the DataKMU network regionally, nationally, and internationally. One of the next steps of the consortium is to apply for a follow-up grant in the realm of ecologically-oriented data science for businesses and industries, where several of the above-mentioned strategic benefits will again play an important role, which will also be in line with the European Green Deal initiative. We look forward to presenting new findings at one of the next iDSC conferences.

Thomas J. Heistracher
Department Head Informatics and Software Engineering
Research Director Information Technologies
Salzburg University of Applied Sciences | Fachhochschule Salzburg GmbH

Organization

Organizing Institutions
AIT - Austrian Institute of Technology
Salzburg University of Applied Sciences

Conference Founders and General Chairs
Manfred Mayr
Peter Haber
Thomas Lampoltshammer

Local Conference Chairs
Helmut Leopold
King Ross
Michael Mürling

Organising Committee
Athina Lykou
Gabriela Viale Pereira
Helmut Leopold
Julian Nöbauer
Kathrin Plankensteiner
Manfred Mayr
Maximilian Tschuchnig
Michael Mürling
Peter Haber
Robert Merz
Thomas Lampoltshammer
Valerie Albrecht