

Editorial

Daniel Baier, Andreas Geyer-Schulz, Berthold Lausen, Angela Montanari,
Józef Pociecha

This issue contains revised versions of selected application-oriented papers presented during

- the sixth European Conference on Data Analysis (ECDA) that took place March 18-20, 2019, at the University of Bayreuth, Germany, and
- the fifth German-Polish Seminar on Data Analysis and Applications (GPS-DAA) that took place at the same location, one day before.

Daniel Baier

Department of Law, Business & Economics, University of Bayreuth, 95447 Bayreuth, Germany

✉ daniel.baier@uni-bayreuth.de

Andreas Geyer-Schulz

Department of Economics & Business Engineering, KIT, 76131 Karlsruhe, Germany

✉ andreas.geyer-schulz@kit.edu

Berthold Lausen

Department of Mathematical Sciences, University of Essex, Colchester CO4 3SQ, UK

✉ blausen@essex.ac.uk

Angela Montanari

Department of Statistical Sciences, University of Bologna, 40126 Bologna, Italy

✉ angela.montanari@unibo.it

Józef Pociecha

Department of Statistics, Cracow University of Economics, 31-510 Cracow, Poland

✉ jozef.pociecha@uek.krakow.pl

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST)

KIT SCIENTIFIC PUBLISHING

Vol. 6, No. 2, 2020

DOI 10.5445/KSP/1000098012/00

ISSN 2363-9881



ECDA in Bayreuth (2019) followed previous editions in Luxembourg (2013), Bremen (2014), Colchester (2015), Wroclaw (2017), Paderborn (2018) and was jointly organized by the University of Bayreuth and

- the Gesellschaft für Klassifikation (GfKl) Data Science Society as well as
- the British Classification Society (BCS),
- the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG),
- the European Association for Data Science (EuADS),
- the Japanese Classification Society (JCS), and
- the Section on Classification and Data Analysis of the Polish Statistical Association (SKAD),

In addition to ten plenary and semi-plenary talks, numerous sessions and lectures (altogether 160 presentations) took place on topics like Clustering, Classification, Data Science, Image Analysis and Computer Vision, Social Network Analysis, Symbolic Data Analysis, and applications of these methods in Biostatistics, Marketing, Medicine and Health Care, and Social Sciences. Also, specialized sessions were thankfully organized by colleagues, e.g.,

- Bioinformatics and Statistics (organized by Dominik Heider),
- Complexity, Data Science and Statistics Through Visualization and Classification (Carmela Iorio, Roberta Siciliano),
- Consumer Preferences and Marketing Analytics (Reinhold Decker, Winfried Steiner),
- Data Analysis in Finance (Krzysztof Jajuga, Herrmann Locarek-Junge),
- Data Analysis Models in Economics and Business (Józef Pociecha),
- Interpretable Machine Learning (Johannes Fürnkranz, Eneldo Loza Mencía, Ute Schmid),
- Statistical Learning (Angela Montanari).

The venue was the main campus of the University of Bayreuth. Besides the talks and discussions also for physical well-being and sightseeing of the 240 participants was cared for: There was a city tour to the Hofgarten with the new and the old castle, to the UNESCO World Heritage listed opera house and to the baroque Friedrichstrasse. A welcome reception was given in the Wagner Museum after a visit of the residence of Richard Wagner, the Villa Wahnfried. The conference dinner took place at Maisel's brewery. Daniel Baier and his local organization team (with many thanks especially to Karolina Ewers and

Benedikt Brand), the 45-member program committee, and the organizers of the specialized sessions have delivered a great event. GPSDAA followed previous German-Polish seminars in Aachen (2009), Krakow (2011), Dresden (2013), Wroclaw (2017), consisted of 13 presentations and was organized by Daniel Baier and Józef Pociecha.

As already mentioned this issue (issue 2 of this Archives of Data Science volume) contains revised versions of selected application-oriented papers of these two conferences. Other papers – with a stronger methodological focus – are contained in another issue (issue 1 of this Archives of Data Science volume) and in a special issue on “Learning in Data Science: Theory, Methods and Applications” of the Springer journal *Advances in Data Analysis and Classification* (ADAC). This way too we wish to thank the 45-member program committee and the organizers of the specialized sessions for supporting us in the selection and reviewing of the many presentations and submitted papers for these three special issues. The 15 selected and revised papers contained in this issue have an application-oriented focus.

The issue starts with 11 papers on **applications of data science methods in marketing**: Carsten D. Schultz discusses whether weekly amounts of categorized search data (using Google Trends) can be used to improve weekly sales forecasts for durable goods (sales of audio books, blue-ray player, TV and so on based on GfK household panel data). Timo Schreiner, Alexandra Rese, and Daniel Baier discuss new developments at recommender systems for online shops and which improvements consumers prefer, based on a large-scale conjoint study. Björn Stoecker, Aydin Nasserli also analyze customers’ online shops preferences based on a large-scale study, but they use a penalty reward contrast analysis for this purpose. Tobias Albrecht and Daniel Baier apply deep neural networks for predicting which customers have a high churn probability at a major mobile phone service provider. Bastian Werner, Ines Brusch, and Larissa Bürks-Arndt discuss which pricing of different product versions satisfies customers and – at the same time – retains profit. Akinori Okada and Hiroyuki Tsurumi apply new kinds of brand switching matrices to a sample of 47,633 consumers buying potato snacks in a time period of two months. They show that the new approach supports in gaining new insights in brand competition. Thomas Reichstein, Ines Brusch, and Rebecca Meier zu Ummeln discuss which announcement formats for events have best chances to go viral in the social networks. They perform an interesting experiment and show that emotional images had the strongest impact across all experimental factors. Evangelia N. Markaki and Theodore Chadjipantelis also discuss the impact of

social media in communication. They perform a conjoint study with 473 respondents to derive important factors. Atsuho Nakayama, Adriana Paliwoda Matiolanska, and Emilia Smolak-Lozano apply text mining to find out which trends companies can derive from Twitter data. The focus lies on sustainability trends in the energy sector. Karolina Ewers and Daniel Baier also discuss new communication forms in marketing: The technology acceptance model and the uses and gratifications approach is applied to measure digital voice assistant (DVA) acceptance in company customer communication. The results are promising: DVAs seem to be an interesting new tool for companies when contacting their customers, the two approaches provided similar results in this direction. Cristopher Siegfried Kopplin and Daniel Baier also discuss the usage of modern software tools in companies: A choice-based conjoint study is used to measure preferences of users with respect to modern workstream collaboration systems like Slack or Trello. A new approach is proposed which helps to develop an ideal support for the users.

Concerning the topic **applications of data science methods in traffic analysis** the paper by Katherina Meißner and Julia Rieck discusses how frequent item sets and rules can be derived from a publically available database of road accident data across six years in UK: The occurrence of these frequent item sets over time and their comparison with the occurrence of accidents form the bases for a decision support system that should allow police analysts to predict critical situations. Two papers follow with **applications of data science methods in finance**: Riccardo Lucato, Edgar Jimenez, Eduardo Salvador Rocha, Yang Qi, Marina Gavriluk, and Rafael Rêgo Drumond discuss ideas for algorithmic trading of stocks at the stock market using LSTM deep learning approaches and sliding windows. The results look promising for investors. On the other side, Katarzyna Kuziak and Krzysztof Piontek analyze systemic risks and risk spill-overs in the Polish financial system. They develop measures that – hopefully – could prevent the markets from the next financial crises. The issue closes with an **applications of data science methods in journalism**. Claus Weihs, Marcel Pauly, and Patrick Stotz argue that journalists should use advanced statistical methods in their articles and demonstrate that this methods could prevent readers from misunderstanding.

Bayreuth, March 2020

Daniel Baier, Andreas Geyer-Schulz, Berthold Lausen, Angela Montanari, Józef Pociecha