



Special issue on “Machine learning and databases”

Matthias Boehm¹ · Nesime Tatbul²

Accepted: 11 March 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Recent advances in machine learning (ML) techniques have led to an explosion in their adoption across all fields of computer science, including database (DB) systems and data management. Meanwhile, end-to-end ML pipelines built for diverse data-driven applications are becoming increasingly more data-centric, presenting new challenges and opportunities in data science and engineering. From data collection and preparation to model training and deployment, efficient access to high-quality data and models form a critical component of the iterative lifecycle of these pipelines. Furthermore, new synergies arise in applying ML techniques for improving DB system internals as well as specializing their functionality and performance to new data and query workload characteristics—a critical need in increasingly more complex deployment environments, such as disaggregated cloud data centers or heterogeneous hardware settings.

This special issue combines innovative research articles that explore data management problems spanning broadly across machine learning and databases (both ML techniques for addressing challenges in data management systems and applications as well as DB techniques for addressing challenges in machine learning systems and applications). We received a total of 18 submissions, out of which 13 have been accepted after a rigorous reviewing process that involved between 1 and 3 revision cycles per paper. 5 of these accepted

articles were brand-new submissions to the journal, whereas the remaining 8 were extended versions of previously published conference papers (4 from SIGMOD, 3 from PVLDB, and 1 from SIGSPATIAL). Overall, our special issue brings together a diverse collection of recent research contributions that lie in the intersection of machine learning and databases, covering a variety of topics including: (i) ML for DB (incremental view maintenance, database performance tuning, spatial join query optimization, adaptive indexing, entity matching, semantic schema alignment, interactive data exploration, synthesizing SQL queries from natural language questions), (ii) DB for ML (efficient linear algebra computations over relational data, reliability evaluation of predictions for trustworthy AI, human-in-the-loop design of data science pipelines, automated optimization of ML pipelines).

We thank the authors for their valuable contributions and the reviewers for their careful assessments, which we believe made this issue a truly special one. By providing a current snapshot of the state of the art in this fast-moving research area, we hope that our special issue can serve as a seminal reference for the readers going forward.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

As the guest editors of the Special Issue on “Machine Learning and Databases”, we are excited to present to the readers of the VLDB Journal an interesting collection of papers from this highly active area of research.

✉ Nesime Tatbul
tatbul@csail.mit.edu

Matthias Boehm
matthias.boehm@tu-berlin.de

¹ TU Berlin, Berlin, Germany

² Intel Labs and MIT, Cambridge, MA, USA