

Madelon Hulsebos

Amsterdam, Netherlands | madelon.hulsebos@gmail.com | madelonhulsebos.com

I do research on the intersection of **machine learning and data management**, with my current focus on **table representation learning** and, more broadly, **intelligent data systems** for insight retrieval.

EDUCATION

University of Amsterdam – Amsterdam, Netherlands
PhD Computer Science Sep 2020 – Feb 2024

Delft University of Technology – Delft, Netherlands
MSc Computer Science Sep 2016 – July 2018
BSc Technology, Policy and Management Sep 2011 – July 2015

POSITIONS

University of California, Berkeley – Berkeley, United States
Postdoctoral Scholar, EECS and BIDS Nov 2023 – present

- Working with Prof. Aditya Parameswaran at the EPIC Lab.

University of Amsterdam – Amsterdam, Netherlands
PhD Researcher, INDE Lab Jan 2023 – Nov 2023
Guest Researcher, INDE Lab Aug 2020 – Jan 2023

- Developed and pursued a research agenda on learned table representations and applications.
- Advised by Prof. Paul Groth.

Sigma Computing – San Francisco, United States
PhD Student Researcher, previously Research Intern (Summer 2021) June 2021 – Dec 2022

- Developed a system for learned table models, and applications like data cleaning and search.
- Contributed to the design and implementation of an interactive ML tool (Decision Studio).
- Published and presented the results at scientific conferences, e.g. CIDR '22.

KPN/HEINEKEN – Rotterdam/Amsterdam, Netherlands
Data Scientist Mar 2019 – May 2021

- Built ML tools for financial forecasting and marketing analyses using e.g. Bayesian models.
- Mentored data analysts, initiated a reading group and process for continuous feedback.
- Gave tutorials and talks on, e.g., data validation and transfer learning.

Massachusetts Institute of Technology – Cambridge, United States
Visiting Collaborator, MIT Media Lab Aug 2018 – Mar 2019

- Led a research project on learned type detection in tables (Sherlock). Sherlock is frequently used in industry and research, and is among the 5 most popular GitHub repos of the MIT Media Lab.
- Contributed to a data visualization training dataset and benchmarking project (VizNet).
- Supervised by Kevin Hu, hosted by Prof. César Hidalgo.

Delft University of Technology – Delft, Netherlands
Graduate TA, Pattern Recognition & Web Information Systems groups Sep 2017 – Feb 2018

- TA for the MSc courses: Pattern Recognition (IN4085), Web Science & Engineering (IN4252).
- Supported 250+ graduate students in labs and projects, and evaluated student assignments.

Aalto University – Helsinki, Finland
Research and Teaching Assistant, Machine Learning for Big Data July - Oct 2017

- Developed material for a BSc course on Machine Learning, in which 500+ students participated.
- Conducted experiments for semi-supervised learning over networks, and presented at ICASSP.

BOARD MEMBERSHIPS (PRO BONO)

UniPartners Delft – Delft, Netherlands

Supervisory Board Member

May 2017 – Dec 2023

- Supervised the financial and strategic position of the student consulting company, and implemented structured financial control, KPI oriented leadership and a supervision cycle.
- Consulted on software projects.

Executive Board Member

Feb 2015 – Feb 2016

- Controlled and optimized the quality of products & processes, and moderated the CRM system.
- Daily management of projects, contributing to a revenue of over €100K.

ACADEMIC SERVICE

Editorship

Assistant Editor, Journal of Systems Research (JSys)

2022 - 2023

Organizing CommitteesCo-organizer, Data Management for End-to-End ML workshop (DEEM) @ **SIGMOD** 2023, 2024Steering Committee, Tabular Data Analysis workshop (TaDA) @ **VLDB** 2023, 2024Founder and co-organizer, Table Representation Learning workshop (TRL) @ **NeurIPS** 2022, 2023Co-organizer, SemTab challenge @ **ISWC** 2021 - 2023**Program Committees**

PVLDB 2024, 2025

ICLR Workshop Proposals 2024

ICDE (Industry track) 2024

aiDM Workshop @ SIGMOD 2024

Data-centric Machine Learning Research Workshop @ ICML 2023, 2024

DBML Workshop @ ICDE 2023, 2024

PhD Workshop @ VLDB 2023

EDBT (Industry track) 2022, 2023

TheWebConf (Industry track) 2022, 2023

NeurIPS (Datasets & Benchmarks track) 2021 - 2023

SemTab @ ISWC 2021 - 2023

AIDB Workshop @ VLDB 2022

University Committees

Award Selection Committee, Amsterdam AI Thesis Awards 2023

ADVISING

T. Cong, PhD dissertation committee member, University of Michigan 2024

W. Lin, MSc research advisor @ UC Berkeley, Semantic Dataset Search 2024

T. Mathijssen, MSc thesis advisor @ UvA, Distributed Query Processing in Apache Calcite 2023

M. Margaret, MSc thesis examiner @ UvA, Leveraging Table Semantics for Data Discovery 2022

INVITED TALKS

Towards Table Representation Learning for end-to-end data management and analysis

Transformers at Work 2023, Zeta Alpha, Amsterdam, Netherlands Sep 2023

INRIA-Saclay, Paris, France Apr 2023

ML for Systems and Systems for ML Workshop @ BTW, Dresden, Germany Mar 2023

Hasso Plattner Institute, Berlin, Germany Mar 2023

KomPAKI seminar, TU Darmstadt, Darmstadt, Germany June 2022

GitTables: a large corpus of relational tables

Tabular Data Analysis workshop, VLDB, Vancouver, Canada
Database Architectures group, CWI, Amsterdam, Netherlands

Aug 2023
Feb 2022

NOTABLE

Postdoctoral Fellowship, Berkeley Institute for Data Science - Accenture	2023
Best Reviewer Award, PhD Workshop, VLDB	2023
Travel Award, VLDB Endowment	2022
Honorable mention GitTables, SemTab challenge	2021

PUBLICATIONS

2024

SPADE: Synthesizing Assertions for Large Language Model Pipelines, **Under Review**

Shankar, S., Li, H., Asawa, P., Hulsebos, M., Lin, Y., Zamfirescu-Pereira, J., Chase, H., Fu-Hinthorn, W., Parameswaran, A., Wu, E.

2023

AdaTyper: Adaptive Semantic Type Detection, **Under Review**

Hulsebos, M., Groth, P., Demiralp, C.

Introducing the Observatory Library for End-to-End Table Embedding Inference, **TRL @ NeurIPS**
Cong, T., Sun., Z., Groth, P., Jagadish, H., Hulsebos, M.

Observatory: Characterizing Embeddings of Relational Tables, **Proceedings of VLDB**

Cong, T., Hulsebos, M., Sun., Z. Groth, P., Jagadish, H.

GitTables: A Large-Scale Corpus of Relational Tables, **SIGMOD**

Hulsebos, M., Demiralp, C., Groth, P.

Models and Practice of Neural Table Representations [tutorial], **SIGMOD**

Hulsebos, M., Deng, X., Sun, H., Papotti, P.

Seventh Workshop on Data Management for End-to-End Machine Learning (DEEM), **SIGMOD**

Boehm, M., Hulsebos, M., Shankar, S., Varma, P.

2022

GitSchemas: A Dataset for Automating Relational Data Preparation Tasks, **DBML @ ICDE**

Döhmen, T., Hulsebos, M., Beecks, C., Schelter, S.

Making Table Understanding Work in Practice [abstract], **CIDR**

Hulsebos, M., Gathani, S., Gale, J., Dillig, I., Groth, P., Demiralp, C.

Augmenting Decision Making via Interactive What-If Analysis, **CIDR**

Gathani, S., Hulsebos, M., Gale, J., Haas, P. J., Demiralp, C.

Results of SemTab 2022, **Proceedings of SemTab @ ISWC**

Abdelmageed, N., Chen, J., Cutrona, V., Efthymiou, V., Hassanzadeh, O., Hulsebos, M., Jiménez-Ruiz, E., Sequeda, J. and Srinivas, K.

2021

Results of SemTab 2021, **Proceedings of SemTab @ ISWC**

Cutrona, V., Chen, J., Efthymiou, V., Hassanzadeh, O., Jiménez-Ruiz, E., Sequeda, J., Srinivas, K., Abdelmageed, N., Hulsebos, M., Oliveira, D., Pesquita, C.

2020

Sato: Contextual semantic type detection in tables, **Proceedings of VLDB**

Zhang, D., Suhara, Y., Li, J., Hulsebos, M., Demiralp, C., Tan, W.

2019

Sherlock: A deep learning approach to semantic data type detection, **ACM SIGKDD**

Hulsebos, M., Hu, K., Bakker, M., Zgraggen, E., Satyanarayan, A., Kraska, T., Demiralp, C., Hidalgo, C.

VizNet: Towards a large-scale visualization learning and benchmarking repository, **ACM CHI**

Hu, K., Gaikwad, N., Hulsebos, M., Bakker, M., Zgraggen, E., Hidalgo, C., Kraska, T., Li, G., Satyanarayan, A., Demiralp, C.

2018

The Network Nullspace Property for Compressed Sensing of Big Data Over Networks, **IEEE ICASSP**

Hulsebos, M., Jung, A.