Madelon Hulsebos

Berkeley, USA | madelon@berkeley.edu | madelonhulsebos.com

My research is on the intersection of **machine learning and data management**, with a focus on **representation learning and generative models for relational tables**. The overarching goal of my research is to make insight retrieval from structured data accurate and easy for everyone.

Employment

Centrum Wiskunde & Informatica – Amsterdam, Netherlands
Tenure-track researcher Sep 2024 – present

- Leading a 5-year research lab to work on neural models for insight retrieval from structured data.
- Co-funded by a $1M AiNed Fellowship grant.

University of California, Berkeley – Berkeley, United States
Postdoctoral Scholar, EECS and BIDS Nov 2023 – Aug 2024

- Worked on systems and benchmarks for retrieval over structured data in RAG and data search.

University of Amsterdam – Amsterdam, Netherlands
PhD Researcher, Informatics Institute Jan 2023 – Nov 2023
Guest Researcher, Informatics Institute Aug 2020 – Jan 2023

- Initiated and pursued a research agenda on neural models for structured data, and applications thereof.
- Thesis title: Table Representation Learning, advised by Paul Groth.

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

- Developed a system for adaptive neural table models for applications like data cleaning and search.
- Contributed to the design and implementation of an interactive intelligent ML tool (Decision Studio).

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 research on semantic type detection in tables (Sherlock). Sherlock is well adopted in industry and benchmarking in research. Also contributed to a data visualization benchmarking project (VizNet).
- Advised by Dr. Kevin Hu and 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.

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

Last updated: 2 August, 2024

Board Memberships (pro bono)

UniPartners Delft – Delft, Netherlands

Supervisory Board Member May 2017 – Dec 2023

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

Organizing Committees

Co-chair Tutorial Track @ **VLDB**	2025
Founder and co-organizer, Table Representation Learning workshop (TRL) @ **NeurIPS**	2022 - 2024
Co-organizer, Data Management for End-to-End ML workshop (DEEM) @ **SIGMOD**	2023, 2024
Steering Committee, Tabular Data Analysis workshop (TaDA) @ **VLDB**	2023, 2024
Co-organizer, SemTab challenge @ **ISWC**	2021 - 2023

Program Committees

PVLDB	2024 -2025
aiDM Workshop @ SIGMOD	2024
ICDE (Industry track)	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

Editorship

Assistant Editor, Journal of Systems Research (JSys)	2022 - 2023

Advising

Supervision

R. Lin, MSc thesis advisor, UC Berkeley	2024
R. Xin, MSc thesis advisor, UC Berkeley	2024
W. Lin, MSc research advisor, UC Berkeley	2024
C. Ji, BSc research advisor, UC Berkeley	2024
T. Mathijssen, MSc thesis advisor, University of Amsterdam	2023
M. Margaret, MSc thesis examiner, University of Amsterdam	2022

PhD committees

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

Talks

Table Representation Learning and Retrieval for Structured Data: It's All About Semantics

Microsoft Gray Systems Lab, USA	July 2024

Advances, challenges, and opportunities in Table Representation Learning

University of Washington Seattle, USA	May 2024
Snowflake, USA	May 2024
Google, Sunnyvale, USA	Apr 2024
UC Berkeley, Berkeley, USA	Mar 2024
Transformers at Work 2023, Zeta Alpha, Amsterdam, Netherlands	Sep 2023

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

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
TU Darmstadt, Darmstadt, Germany	June 2022

GitTables: a large corpus of relational tables

Tabular Data Analysis workshop, VLDB, Vancouver, Canada	Aug 2023
Database Architectures group, CWI, Amsterdam, Netherlands	Feb 2022

Awards and Funding

AiNed Fellowship Grant, $993K, NWO	2024
Postdoctoral Fellowship, $150K, Accenture-BIDS	2023
Best Reviewer Award, PhD Workshop, VLDB	2023
Travel Award, $2.5K, VLDB Endowment	2022
Honorable mention GitTables, SemTab challenge	2021

Publications

2024

"It Took Longer than I was Expecting:" Why is Dataset Search Still so Hard?, **HILDA@SIGMOD**
Hulsebos, M., Lin, W., Shankar, S., Parameswaran, A.

Towards Accurate and Efficient Document Analytics with Large Language Models, **Under Review**
Lin, Y., Hulsebos, M., Ma, R., Shankar, S., Zeigham, S., Parameswaran, A. G., & Wu, E.

SchemaPile: A Large Collection of Relational Database Schemas, **SIGMOD**
Doehmen, T., Geacu, R., Hulsebos, M., Schelter, S.

SPADE: Synthesizing Assertions for Large Language Model Pipelines, **Proceedings of VLDB**
Shankar, S., Li, H., Asawa, P., Hulsebos, M., Lin, Y., Zamfirescu-Pereira, J., Chase, H., Fu-Hinthorn, W., Parameswaran, A., Wu, E.

Eighth Workshop on Data Management for End-to-End Machine Learning (DEEM), **SIGMOD**
Hulsebos, M., Shankar, S., Interlandi, M.

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.V.

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

GitTables: A Large-Scale Corpus of Relational Tables, **SIGMOD**
Hulsebos, M., Demiralp, C., Groth, P.

GitSchemas: A Dataset for Automating Relational Data Preparation Tasks, **DBML @ ICDE**
Döhmen, T., Hulsebos, M., Beecks, C., Schelter, S.

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.

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.

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.

Last updated: 2 August, 2024