The Michigan Institute for Data Science (MIDAS) empowers innovative, data-driven research, leveraging U-M faculty expertise across disciplines.

MIDAS is part of Advanced Research Computing (ARC) at the University of Michigan. ARC provides a comprehensive ecosystem of resources and support for researchers engaged in data-intensive and computational science. ARC consists of MIDAS, the Michigan Institute for Computational Discovery and Engineering (MICDE), Advanced Research Computing - Technology Services (ARC-TS), and Consulting for Statistics, Computing and Analytics Research (CSCAR).

Cover illustration: The Michigan Data Science Team, supported by MIDAS, predicted the location of pipes with the greatest chances of needing replacement during Flint’s water crisis.
The power of data science to solve humanity’s most vexing problems is on full display in the research being done by the students and faculty supported by the Michigan Institute for Data Science.

From the lead-contaminated water pipes of Flint, Michigan to the emerging fields of automated vehicles, precision health, and single-cell biology, and beyond, MIDAS researchers push scientific knowledge forward, in service of the public good.

Since its founding in 2015, MIDAS has supported a wide range of research initiatives, investing more than $10 million in interdisciplinary projects and leveraging the expertise of U-M’s world-class faculty and students.

We also foster emerging research ideas by convening working groups of faculty members engaged in similar areas of inquiry in different academic departments. This collaborative, cross-cutting approach is resulting in projects that defy traditional categories, but harness the power of data science in new and effective ways.

MIDAS is cultivating the next generation of data scientists through a myriad of educational opportunities including a graduate certificate in data science, a multitude of online courses, and a summer camp for high school students interested in big data.

In partnership with several Michigan companies, MIDAS researchers are helping industry realize the potential of data science in medicine, finance and transportation.

We invite you to peruse some of our recent accomplishments in this report. See midas.umich.edu to learn more about us and to contact us for potential collaborative opportunities.

Core MIDAS faculty members: Building the data science community at U-M
Data-driven solutions for the Flint water crisis

The Michigan Data Science Team (MDST) and its faculty advisors Eric Schwartz and Jacob Abernethy have been helping citizens of Flint, Mich., predict where lead-contaminated pipes may be as the city of nearly 100,000 struggles with its years-long water safety crisis. Using data from digitized versions of decades-old city records in combination with more recently produced datasets, the team’s predictive models could save the city as much as $10 million as it replaces pipes containing lead.

A paper describing the research, co-authored by Abernethy, Schwartz, Alex Chojnacki, Arya Farahi and Jared Webb, titled “Active Remediation: The Search for Lead Pipes in Flint,” won the Best Student Paper award at the KDD Conference in London in August, 2018. DOI: 10.1145/3219819.3219896

Single-cell analysis yields progress on knowledge of sperm production

U-M researchers used single-cell analysis of 30,000 cells from mouse testes to improve the understanding of spermatogenesis. The research identified new types of cells and molecular features in sperm cell development. Lead researchers Sue Hammoud, PhD, Assistant Professor of Human Genetics, Obstetrics and Gynecology, and Urology; and Jun Li, PhD, Professor of Human Genetics, and Computational Medicine and Bioinformatics, are part of the MIDAS-supported Michigan Center for Single-Cell Genomic Data Analytics.

The findings were published in Developmental Cell. DOI: 10.1016/j.devcel.2018.07.025

Fake news detector works better than humans

An algorithm-based system that identifies telltale linguistic cues in fake news stories could provide news aggregator and social media sites like Google News with a new weapon in the fight against misinformation. Rada Micalcea, Ph.D., Professor of Electrical Engineering and Computer Science, developed the system and has demonstrated that it’s comparable to and sometimes better than humans at correctly identifying fake news stories.

The research was supported by MIDAS and the NSF, and presented to the 27th International Conference on Computational Linguistics. URL: https://arxiv.org/pdf/1708.07104.pdf

MIDAS funds Data Science for Music projects

From digital analysis of Bach sonatas to mining data from crowdsourced compositions, U-M researchers are using modern big data techniques to transform how we understand, create and interact with music. Four U-M research teams have received support for projects that apply data science tools like machine learning and data mining to the study of music theory, performance, social media-based music making, and the connection between words and music.
MIDAS Research Hubs

MIDAS has invested more than $10 million in cross-cutting research that takes advantage of emerging data science methodologies and applications. Our research hubs leverage these investments by fostering collaboration between faculty and students in a variety of disciplines focused on some of the most important problems in our society:

- **Transportation: The Next Generation of Mobility**
  - Reinventing Public Urban Transportation and Mobility
  - Building a Transportation Data Ecosystem

- **Health Science: Advancing the Spectrum of Biomedical Research**
  - Michigan Center for Single-Cell Genomic Data Analytics
  - Michigan Integrated Center for Health Analytics and Medical Prediction
  - Identifying Real-Time Data Predictors of Stress and Depression using Mobile Technology

- **Learning Analytics: Improving Education through Data**
  - Analytics for Learners as People
  - Holistic Modeling of Education

- **Social Science: Leading the Evolution of Social Research**
  - Computational Approaches for the Construction of Novel Macroeconomic Data
  - A Social Science Collaboration for Research on Communication and Learning Based Upon Big Data

- **Data Science for Music**
  - Understanding and Mining Patterns of Audience Engagement and Creative Collaboration in Large-Scale Crowdsourced Music Performances
  - Understanding How the Brain Processes Music Through the Bach Trio Sonatas
  - The Sound of Text
  - A Computational Study of Patterned Melodic Structures Across Musical Culture

MIDAS Research Working Groups

To plant the seeds of tomorrow’s data science innovations, MIDAS has convened several cross-disciplinary research working groups. The groups are either focused on specific funding opportunities, or themes that cut across traditional disciplines. The goal is to foster interactions between theorist and application scientists, enable innovative ideas and new collaboration, and elevate the quality of data science research across campus. The current groups are:

- Data Integration
- Data Science for Music
- Mobile Sensor Analytics
- Teaching Data Science
- Trustworthy Data Science

For more on MIDAS Research Hubs and Working Groups, see midas.umich.edu/research.
New offerings strengthen U-M’s data science education program

The job outlook for data scientists is one of the most promising in any field. The U.S. Labor Department predicts a 20 percent increase in the next decade. To meet this demand, U-M and MIDAS have been expanding educational opportunities for students of all levels and skill sets.

MIDAS offers the Graduate Certificate in Data Science, which emphasizes analysis methods, data management, and algorithms and applications, has nearly 100 enrolled students.

The Masters of Science in Data Science, created in 2018, is a joint program offered through the departments of Electrical Engineering and Computer Science, Statistics, and Biostatistics, the School of Information, and MIDAS. The Masters of Applied Data Science is an online-only degree offered in partnership with Coursera.

See midas.umich.edu/education for an outline of existing programs.

Summer camp introduces data science to high school students

For the last three years, MIDAS has run a summer camp that has introduced the concepts and practice of data science to high school students from Michigan and across the country. Participants have the opportunity to speak with U-M researchers and representatives of data-driven companies and government organizations; experience museums, sports teams, and other local attractions through a data-science lens; and learn about real-world applications of data science tools.

For more information, visit midas.umich.edu/camp.

MIDAS-supported student groups

MIDAS is proud to support student groups with a combined active membership of more than 400 students and 50 faculty members.
MIDAS expands corporate partnerships

MIDAS is pleased to work with corporate partners on research projects; sponsorships of seminars and other events; recruitment and career events; and other initiatives. Please email us at midas-contact@umich.edu, for more information about engaging with MIDAS.

MIDAS, Peers Health partner on disability and workers comp study

Peers Health and MIDAS have begun a two-year research project that will apply advanced learning technologies to a proprietary global database of millions of de-identified disability and workers’ compensation cases.

The goals of the project include developing a prescriptive modeling framework to facilitate development of optimal return-to-work plans for injured or ill patients.

The findings from this project, by helping define when someone objectively has returned to health, could inform decision-making in virtually every healthcare episode.

The principal investigators in the project are Dr. Brian Denton, Industrial and Operations Engineering and Urology, and Dr. Jenna Wiens, Computer Science and Engineering, both of the University of Michigan.

Students take part in Quicken Lending Strategies Prediction Challenge

U-M students in the Michigan Data Science Team took part in a data science challenge sponsored by Quicken Loans. The goal was to create a model that would predict whether potential clients would end up getting a mortgage based on the loan product originally offered to them.

Students were given access to proprietary, de-identified financial data from recent Quicken clients, and the accuracy of the models was evaluated based on one month of client data.

The winners presented their work to the Quicken Data Science team at the company’s headquarters in downtown Detroit.
MIDAS by the numbers

- 200+ core and affiliate faculty members from Ann Arbor, Dearborn and Flint campuses
- 5 data science research hubs: transportation, learning analytics, health science, social science and music
- 13 corporate partners engaged in sponsored research, talent recruitment and event sponsorship
- 52 graduate students in 12 degree programs have earned MIDAS Data Science Rackham certificates. More than 110 are enrolled or applying.
- 400+ students and 50+ faculty members take part in MIDAS-sponsored data science student organizations
- 14 public service projects across Southeast Michigan have been completed by MIDAS supported data science student organizations
- 70+ events (symposia, workshops, seminars) building the data science community at U-M

midas.umich.edu

The Michigan Institute for Data Science is a unit of the University of Michigan Office of Research.
The University of Michigan is a non-discriminatory, Affirmative Action Employer.
© 2018 by the Regents of the University of Michigan.