UA Institutes Work to Build Data Literacy and Transdisciplinary Research

The Office of Research, Discovery and Innovation
December 2018

Two new institutes have put the UA on track to become a data science powerhouse at the national level and leader in promoting data science literacy and transdisciplinary research.

The UA's Data Science Institute, known as Data7, a unit of the **Office of Research, Discovery and Innovation**[1], is focused on connecting researchers, fostering collaboration and promoting literacy across campus. **UA-TRIPODS** [2], an integrated research and educational institute funded by the National Science Foundation, shares these goals and also is focused on developing new algorithms and foundational approaches necessary for large-scale data-driven research.

Nirav Merchant [3], Data7's director, said he and **Hao Helen Zhang** [4], mathematics professor and principal investigator for UA-TRIPODS, "are working to position data science at the nexus of the UA's cutting-edge research."

"Our goals are to advance the field of data science, facilitate interdisciplinary collaboration, enhance graduate and undergraduate literacy, promote workforce preparation, and foster industry alliances," Merchant added.

One of the first steps to accomplishing these goals, he said, is to promote data literacy across campus.

To that end, Data7 partnered with the Data Science Resources and Training Steering Committee to develop the **Data Science Ambassadors Program** [5], an effort to bring data science to more people across campus. The **steering committee** [6], made up of researchers and data science experts from across the UA, advocates, advises and cultivates new data science opportunities at the UA.

Data literacy = building community + building capacity

All researchers need a basic level of data literacy, including understanding how to work with data, think with data and manipulate data, said **Vignesh Subbian** [7] and **Jeffrey Oliver** [8], co-directors of the Data Science Ambassadors Program. Then, they need to understand how to make decisions and frame new questions based on the data.

"The way to enhance data literacy on campus is to build community and build capacity," said Subbian, an assistant professor of biomedical engineering and of systems and industrial engineering, and a member of the BIO5 Institute.

Building community involves connecting experts in various fields on campus with colleagues who are experts in data science, said Oliver, a data science specialist with the **Office of Digital Innovation and Stewardship**[9] at University Libraries. Building capacity, he added, entails "deepening data science expertise within specific departments and among individual researchers."

Launched earlier this fall, the program now has graduate student ambassadors in the **College of Social and Behavioral Sciences** [10], the **College of Engineering** [11], the **College of Science** [12], and the **College of Agriculture and Life Sciences** [13].

"Each ambassador has both data science and domain expertise and acts as a node in the web of community connections," Oliver said.

The ambassadors' role, Subbian added, is to help researchers in their respective colleges – including students – with data science-related questions through consulting, training or referrals to resources or services on campus.

"There are a lot of silos in academia," Oliver said. "We believe ambassadors with one foot in Data7 and one foot in their respective colleges can do a lot to break through artificial barriers and help catalyze data science efforts across our campus."

A catalyst for collaboration

Sometimes revolutionary scientific discoveries are made by serendipitous accidents. But, as associate professor of computer science **Mihai Surdeanu** [14] points out, "relying on chance is no way to do science."

Although Surdeanu has been doing some fascinating science based on his own serendipitous connections, he is now actively promoting Data7 as a better way to connect with like-minded researchers and catalyze otherwise unlikely collaborations.

One of Surdeanu's own unlikely collaborations started in 2013 when Surdeanu happened to hear a radio interview with **Melanie Hingle** [15], assistant professor of nutritional sciences who studies ways to predict and avoid obesity and Type 2 diabetes, particularly among children and families. Hingle was being interviewed about the work she has done with data visualization expert **Stephen Kobourov** [16], a professor of computer science and now Data7's associate director, and one

of Hingle's undergraduate students. The team used tweets to understand whether social media users' posts about food could predict their risk of developing diabetes.

"It just clicked," Surdeanu said as he recalled hearing the interview. "I thought, 'We could actually predict all sorts of health outcomes if we had this kind of data collected at scale.'"

Soon after, Surdeanu contacted Hingle and wrote a computer program, called a bot, that searched Twitter for #breakfast, #dinner, #lunch and other food-related hashtags to find tweets about food. The bot's been running for five years and counting, and has collected 25 million tweets.

"Using machine learning, we train a classifier – a computer model that learns automatically – to sort through the data set, and then we predict the risk level for obesity or diabetes for an individual or for the population of a geographic area," Surdeanu said.

The model's results are then validated by asking individuals to complete risk surveys and by comparing them with existing information from the Centers for Disease Control and Prevention.

For Hingle, the collaboration has been particularly rewarding.

"There are 85 million people in the U.S. alone who are prediabetic and don't know it," she said. "The idea that we can use computer science techniques and social media to make accurate predictions about disease risk is very exciting."

Since starting its work together, the team has expanded to include **Stephen Rains** [17], professor of communication. Now, the group is trying to find ways to deploy its machine-learning tool with messaging that will motivate people to make better food choices to reduce their disease risk, and do so without raising privacy concerns.

Make the connection

Data7's mission to encourage interdisciplinary connections and foster data literacy extends throughout the UA community, Merchant said, adding that affiliates in all departments with research interests in machine learning, natural language processing, image analysis, large-scale visualization or literacy are encouraged to join as capabilities and partnerships are developed.

For additional information on the Data Science Institute and the ambassadors program, visit the institute's **website** [18]. For information on high-performance computing services, click **here** [19].

Source URL:https://uaatwork.arizona.edu/lqp/ua-institutes-work-build-data-literacy-and-transdisciplinary-research

Links
[1] https://research.arizona.edu/ [2] https://tripods.math.arizona.edu/ [3] https://datascience.arizona.edu/person/nirav-merchant [4] https://math.arizona.edu/people/haozhang [5] http://ceero.collab.arizona.edu/dsa.html [6] https://datascience.arizona.edu/steering-committee [7] https://sie.engineering.arizona.edu/faculty-staff/faculty/vignesh-subbian [8] https://new.library.arizona.edu/people/jeffrey-oliver-phd [9] https://new.library.arizona.edu/departments/odis [10] https://sbs.arizona.edu/ [11] https://engineering.arizona.edu/ [12] http://cos.arizona.edu/ [13] https://cals.arizona.edu/ [14] http://uacc.arizona.edu/profile/mihai-surdeanu [15] https://nutrition.cals.arizona.edu/people/melanie-hingle-phd-mph-rd [16] https://www2.cs.arizona.edu/~kobourov/ [17] https://comm.arizona.edu/user/stephen-rains [18] https://datascience.arizona.edu/ [19] https://it.arizona.edu/service/high-performance-computing