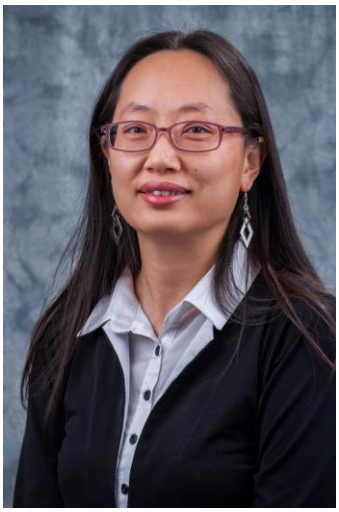
Texas A&M Institute of Data Science Tutorial Series

TEXAS A&M Institute of Data Science

Iterating Between the Data World and the Real World to Find Answers in Big Data



"*Data science is the systematic study of digital data using scientific techniques of observation, theory development, systematic analysis, hypothesis testing, and rigorous validation*" (Ahalt, 2012). Data scientists are those that can apply data science to continuously changing deluge of digital raw data that are often inconsistent and erroneous to extract and deliver actionable knowledge in a timely manner. Data scientists are interdisciplinary scientists who have a combination of skills in statistics and data mining (to turn raw data into information), programming (to build the data pipeline and infrastructure for efficient processing of raw data into information), and domain expertise (to correctly translate between reality and data, know what information is required, and how accurate it needs to be). This tutorial will focus on the critical thinking needed as a data scientist when they iterate between the data world and the real world to find answers in big data projects through multiple case studies from the health sector. We will cover the different trade off decisions that often arise in real projects such as issues of measurement, choice of different modeling methods, and building replicable data pipelines.

The tutorial will be a demonstration that goes through different case studies from healthcare applications, highlighting many tradeoffs in decision-making. After the main tutorial, the speaker will be available to consult on issues of how to handle sensitive person-level data covering topics in IRB, DUA, and computer security.
Background knowledge advisable: Participants need at least a basic knowledge of the data management to develop measures and data modeling.

Hye-Chung Kum, Ph.D.
Professor
Department of Health Policy and Management
Texas A&M University

Date: Friday, October 16
Time: 1:00 – 3:00 p.m. US Central Time
Meeting ID: 998 4499 3279
Passcode: 724615
Faculty host: Yu Ding, TAMIDS

Biography

Dr. Hye-Chung Kum is a Professor in the Department of Health Policy and Management (HPM) with joint appointments in the Departments of Computer Science and Engineering, and Industrial Systems Engineering at Texas A&M University. She is a data scientist cross trained in computer science (PhD in datamining) and Welfare Policy and Management (Master of Social Work) at University of North Carolina at Chapel Hill. For over 20 years, she has been conducting research on methods and tools to use big data about people to answer questions about population health and well-being with a focus on using integrated government administrative data. She is the founder and director of the Population Informatics Lab and have collaborated on diverse areas of research in computer science (record linkage, sequential pattern mining, information security, information privacy, natural language processing (NLP), Human Computer Interaction (HCI), machine learning), health services research, health informatics, welfare policy, nutrition, epidemiology, and ELSI (ethical, legal, and social implications). She has over 50 publications in diverse journals and her research has been funded by NSF, NISOH, PCORI, RWJF, North Carolina Division of Social Services, and Texas Health and Human Services Commission. She is a 2018 presidential impact fellow of the Hagler Institute of Advanced Study and a Royster Fellow.

TEXAS A&M UNIVERSITY | TEXAS A&M AGRILIFE | TEES Texas A&M Engineering Experiment Station

For more information about TAMIDS tutorial series, please contact Ms. Jennifer South at jsouth@tamu.edu