Texas A&M

INSTITUTE OF DATA SCIENCE

Professional Education Workshops in Data Science Foundations and Computational Practice

Data science is a multidisciplinary field that utilizes statistics, data analysis, machine learning, algorithms, software, and computing systems to extract information, acquire knowledge, and gain insights into the underlying context from which data is generated.

The Texas A&M Institute of Data Science (TAMIDS) workshop on Data Science Foundations and Computational Practice is a week-long intensive workshop that equips participants with diverse skills to enable their professional practice of data science.

Target audience: organizations seeking to help their workforce develop knowledge and experience applying computational methods of data science.

Format: a sequence of ten interactive, three-hour modules that integrate methodological exposition with application to data through hands-on programming and computation. Participants will use open source tools and libraries to develop proficiency and understanding of the essential methods of current data science and state-of-the-art topics.

Delivery: the workshop may be delivered remotely in synchronous, asynchronous, mixed modes, or on-site.

Prerequisites: familiarity with Linux or similar operating-system environments, experience with Python or similar programming language, introductory experience with data analysis and statistics.

Instructors: Texas A&M faculty experts deliver the program modules, which were developed and field-tested in data science webinars, boot camps, and tutorials.

Customization: TAMIDS may work by arrangement with other organizations to develop domain-application content for customized workshop events.

LEARNING OUTCOMES

Upon the completion of the course, each participant should be able to:

- Create and manage an open source software environment for data science projects.
- Use open source tools to read, update, and write JSON, CSV, XML, and other structured data formats.
- Apply NumPy and SciPy packages for numerical and statistical computation on data.
- Gain insight into data through analysis and visualization using pandas and matplotlib open source libraries.
- Apply common supervised and unsupervised machine learning methods; identify pitfalls such as over-fitting.
- Design and develop non-trivial programs for data science in Python using libraries and frameworks for machine learning and distributed computation, including scikit-learn and Spark.
- Use reinforcement learning to optimize control of artificial intelligence systems in the absence of a specific reward model.
- Model complex and high-dimensional data by learning latent low-dimensional representations.
- Integrate deep-learning frameworks such as TensorFlow and PyTorch into data analytics workflows.
- Develop models and perform feature selection using automated machine learning.

CONTACT

Dr. Nick Duffield
TAMIDS Director
duffieldng@tamu.edu

Produced by Research Communications 8/2020

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

tamids.tamu.edu