Fall 2014

The Science of Information:
Big Data Analytics and Machine Learning

Tuesdays and Thursdays at 2:00 PM to 3:15 PM

Dr. Shan Suthaharan
University of North Carolina at Greensboro (UNCG)

s_suthah@uncg.edu

The course will focus on modern computer technologies and machine learning techniques, and their usefulness to big data analytics. The topics, in broader terms, will include: defining big data issues; understanding big data technologies; developing machine learning techniques; and solving big data classification problems using machine learning techniques. The course materials will be divided into two parts: Part I-Machine Learning (ML) and Part II-Big Data Analytics.

Students will be introduced to the CSol learninghub site (http://soihub.org) as well as to the references in the site. Students will learn about machine learning approaches (e.g. regression and classification models), Hadoop distributed file system, Scikit-learn ML Libraries (or similar Big Data and Machine Learning tools and associated programming language), and apply ML techniques to various datasets and interpret results.

This course is suitable for senior undergraduate and graduate students in computer science, electrical engineering and mathematics/statistics disciplines. The prerequisites for this course are differential and integral calculus, statistics, and data structures and algorithms (UNCG courses MAT 292, STA 271, and CSC 230). Proficiency in at least one programming language (e.g. Python, C, Java, …) is also expected.

UNCG COURSE CODE CSC495/CSC693

The development and delivery of the course is funded by the Center for the Science of Information, Purdue University through a sub-award approved by the National Science Foundation, and partially funded by UNCG.