

October, 5, 2024

Prof. Raja Sooriamurthi
Information Systems Program
Carnegie Mellon University
4800 Forbes Ave, HBH 3017
Pittsburgh, PA 15213 USA
raja@cmu.edu

Course: Data Analysis and Machine Learning

Course Description:

This course is an introduction to the techniques and tools for analyzing and distilling actionable knowledge from data with the end goal of adding value.

Currently we are in the midst of the next disruptive age of Information Technology. The combined advances in hardware, software, and communication forms the basis of our current disruptive age of data. Organizations and businesses need data driven actionable insights from data. For example, a casino may want to identify whether there is a certain group of customers from which more business occurs—a task known as customer segmentation. A cell phone company may want to know if there is a risk of customers leaving for another carrier—a business situation known as customer churn. Analytic tasks that facilitate such actionable insights include prediction, optimization, recommendation, classification, clustering etc.

This ongoing IT revolution driven by data is also viewed as the fourth paradigm of science. For more than 1000 years science has been driven by empirical methods. Starting a few hundred years ago a mathematics based theoretical science paradigm emerged. As human achievement progressed, it turned out that some phenomena cannot be approached empirically or they are not tractable to theoretical approaches (e.g., earthquakes, thermonuclear fission). Hence, few decades back, yet another paradigm of science fostered computational simulations to study these phenomena. Currently a new paradigm of doing science based on data has emerged—data science. In this course we will study the techniques and tools of these two intertwined themes (i) exploratory data analysis and (ii) machine learning.

Learning Objectives

Upon successful completion of this course, students will be able to:

1. Apply the principles of computational thinking (CT) to data science
2. Demonstrate development practices conforming to the Pythonic way
3. Express a business problem as a data problem
4. Perform exploratory data analysis from inception to the value proposition
5. Explain the core principles behind various analytics tasks such as classification, regression, clustering, recommendation, association.

Working in teams of 3-4, students will learn how to formulate a research question and form a research hypothesis. They will learn how to critically evaluate the research hypothesis and findings, and write up their findings in a final project report (due at the end of the course). They will also learn how to do a thorough literature search.

Evaluation: Yu, Zixuan (Hancy)

Zixuan attended all the courses and lectures, listened carefully in class and always asked questions to the professor. He is a calm and thoughtful student. He always thinks about problems seriously and spends a lot of time on calculations. His stable personality and careful listening to the professors and teaching assistants have benefited him a lot. He is an excellent collaborator and listener. His group has shown research capabilities that exceed the undergraduate level. His coding ability, engineering ability and algorithm ability are all excellent. Overall, he has completed the learning tasks in class and the teamwork tasks of the group with a very positive attitude.

With proper guidance I believe he can do very well in your program. I wish him all success in his academic pursuits.

Please do not hesitate to contact me at my email address above if further information is needed.

Sincerely,



Raja Sooriamurthi

Teaching Professor, Information Systems Program
Program Director, Decision Analytics and Systems (DAS) minor