INFSCI 0510 Data Analysis (Spring 2025)

Instructor: Dr. Yuqi Ouyang
Email: yuqi.ouyang@scupi.cn

Section 1: Wednesday, 8:15 – 11:00, S201
Section 2: Wednesday, 13:50 – 16:25, S201

Office Hours:

Monday, 9:00 – 17:00. Room 524.
Please send an email to schedule a meeting.

Course Description:

This course offers an introductory exploration of data analysis using Python. Participants will learn how to work with data, starting from the selection of appropriate models for diverse tasks. The course progresses to cover essential techniques of pattern analysis using machine learning models.

Course Objectives:

- Machine Learning Context: Understand the machine learning context and the different task scenarios such as classification and clustering.
- Appropriate Models: Gain proficiency in selecting suitable models for various data analysis tasks, understanding the relevance of each model in different contexts.
- Performance Evaluation: Know different evaluation metrics for various data analysis tasks, thus be able to interpret and demonstrate the performance of the model.

Prerequisites:

INFSCI 0310 Computation in Information Science

Textbook:

Mathematics for Machine Learning. Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Cambridge University Press. 2020.

Assessment:

Attendance	10%
Assignments	35%
Project	15%
Final exam	40%

List of Topics:

- Math Preliminaries
- Introduction to Machine Learning
- Supervised Machine Learning:
  (1) Classification: The Perceptron; Logistic Regression; K-Nearest Neighbors; Decision Trees; Naïve Bayes Classifier; Multi-Class Classification; The Kernel Trick; Support Vector Machine.
  (2) Model Evaluation and Selection: The Receiver Operating Characteristic (ROC); Area Under ROC (AUROC); K-fold Cross-Validation.
- Unsupervised Machine Learning:
  (1) Dimensionality Reduction: Principle Component Analysis (PCA)
  (2) Density Estimation and/or Clustering: K-Means Algorithm; K-Means++ Algorithm; Density-Based Spatial Clustering of Applications with Noise (DBSCAN); Maximum Likelihood Estimation (MLE); Maximum A Posteriori (MAP); Gaussian Mixture Models (GMMs); The EM Algorithm.
- Ensemble Methods: Bagging and Boosting.

Mitigating Circumstances:

If you have a medical situation or any personal circumstances that substantially affect your study or exam. You are encouraged to contact the department or the instructor as soon as possible. With valid proof, mitigation may be applied when assessing your assignments, coursework or exam sheets.

Course Policies:

- Please regularly check the announcements on Blackboard
- We can not assure instant respond to emails, we suggest bring urgent questions to face-to-face sessions.
- Google skill is one of the most important skills in Computer Science study, try Google your questions first.
- Treat ChatGPT as an auxiliary tool only. Use your brain first because it is much more powerful.
- Zero-tolerance to both two persons in plagiarism. Checks will be done to detect plagiarism.
- Late submissions cause penalties, unless due to approved mitigating circumstances, prior to the deadline.
- Students with documented emergencies, after careful evaluation, may jump to the make-up exam.
- All course materials or any recordings are for your personal use and for educational purposes only.

Learning Outcomes:

- Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions.
- Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
- Apply computer science theory and software development fundamentals to produce computing solutions.
- An ability to acquire and apply new knowledge as needed, using appropriate learning strategies.