

INFSCI 0510 DATA ANALYSIS (Spring 2024)

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

Section 1: Wednesday, 8:15 – 11:00, 4-212
Section 2: Wednesday, 13:50 – 16:25, 4-212

Office Hours:

Tuesday, 9:00 – 12:00. Thursday, 14:00 – 17:00. New SCUPI building, Room 523.
By appointment only. Please send emails to schedule a meeting.

Teaching Assistants & Lab Hours:

Shiwei Hong (Section 1): sewellhcaulfield@foxmail.com. Friday, 12:00-14:00. 3-105
Hong Liu (Section 2): 2862644418@qq.com. Friday, 12:00-14:00. 3-105

Course Description:

This course offers an introductory exploration of data processing and analysis using the Python programming language. Participants will learn how to work with data, starting from the selection of appropriate models for diverse tasks. The course progresses to cover essential aspects of data munging, followed by pattern analysis using foundational machine learning algorithms and result evaluation.

Course Objectives:

- **Model Selection Proficiency:** Gain proficiency in selecting suitable models for various data analysis tasks, understanding the relevance of each model in different contexts.
- **Data Munging Techniques:** Acquire skills in data munging, focusing on the preprocessing and transformation of raw data to enhance its quality and usability.
- **Pattern Analysis and Evaluation:** Explore the intricacies of pattern analysis using basic machine learning algorithms, and develop the ability to evaluate results effectively.

Learning Outcomes:

- **Model Selection Expertise:** Successfully choose appropriate models for diverse data analysis tasks, understanding the nuances of model selection in specific contexts.
- **Effective Data Munging:** Apply data munging techniques to preprocess and transform raw data, ensuring its quality and suitability for analysis.
- **Pattern Analysis and Evaluation Skills:** Employ basic machine learning algorithms for pattern analysis and demonstrate the ability to critically evaluate and interpret results, contributing to informed decision-making in data-driven scenarios.

Prerequisites:

Students must have studied INFSCI 0310 Computation in Information Science.

Assessment:

Attendance	10%
Assignments	0%
Coursework	40%
Final exam	50%

Tentative Outline:

• Data Analysis Basics:

Introduction, Python Basics, Descriptive Statistics, Data Munging, Data Visualization (Matplotlib)

• Supervised Machine Learning:

Introduction to Machine Learning, Linear Regression, Basic Classification Models, Model Evaluation and Model Selection, Kernel Trick, Support Vector Machine.

• Unsupervised Machine Learning

Basic Clustering Models, Advanced Clustering: Gaussian Mixture Models

• Other Data Analysis Topics:

Principle Component Analysis, Anomaly Detection.

Textbook:

Data Mining: Concepts and Techniques 3rd Edition. Jiawei Han, Micheline Kamber, Jian Pei. 2012.

We suggest focusing on lecture slides and course materials.

Communication and Assistance:

The instructor and TAs may periodically post announcements to the Blackboard website, or email enrolled students with announcements. You should regularly check these announcements. Note that the QQ groups serve for emergency reasons.

When contacting the course staff via email, messages must be addressed to (or CC) both the instructor and the TAs. Email subject should be prefaced with the appropriate prefix (e.g., INFSCI0410).

Regarding providing assistance, we can not assure instant respond to emails. Hence, we suggest bring your questions to face-to-face sessions, e.g., the lab hours.

More importantly, please check the following options, if you are getting an error in your program, what should you first do?

- A. Email your instructor or your TAs every 5 minutes until someone responds.
- B. Bring your errors or questions to face-to-face sessions.
- C. Throw your laptop on the floor and swear to never program again.
- D. Google the error, find a solution quickly from online websites such as StackOverflow.
- E. Use ChatGPT to analyze the error and find a solution.

We recommend option *D*, i.e., Google your error first. When you contact us for help, probably we will also Google your error and get some help from StackOverflow. Hence, why not try it yourself before contacting the course staff? In this case, you will learn a skill for problem-solving, i.e., the Google skill, which is one of the most essential skills in Computer Science study.

Regarding option *E*, we know that ChatGPT has become a powerful tool even for solving programming issues. However, it is not a very good option because ChatGPT outputs words and sentences as the answer that has the highest likelihood to address your question, meaning that ChatGPT will never actually run either the code in your question or the code in the answer. Also, the reliance on ChatGPT is not a good idea because during exams you will not be able to use ChatGPT anymore.

Plagiarism:

Since this module does not contain any group works, students are expected to complete their works independently. Discussions with other students on the assignments or the coursework should be limited to understanding the statement of the problems. Plagiarism will not be tolerated and substantial penalties will be given on both two persons involved in the activity.

Late Submissions:

All submissions should be on time. Late submissions beyond the deadline will not be tolerated and penalties will be given, unless the students' late submissions are due to mitigating circumstances and the late submission has been approved by the instructor, prior to the deadline.

Make-Up Exams:

Students are expected to take the final exam. Make-up exams will only be given in the event of a medical situation or an emergency, and only if this is documented and the instructor is notified immediately. Missing an exam will result in a failure for the exam.

Mitigating Circumstances:

If you have a disability 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.

Audio or Video Recording:

Without the written permission of the instructor, students may record lectures, discussions or activities by audio or video. However, any such recordings should be solely for personal use.