

Course Syllabus

CS5850/CS6850 Introduction to Data Analysis (3 credits)

Spring 202X

Time: —

Location: —

Instructor: Dr. Hamid Karimi

Email: hamid.karimi@usu.edu

Phone: (435) 797-1233 (office)

Office: 418 Old Main Building

Office Hours: –

Delivery Mode: Face-to-face

-

Graduate Teaching Assistant (GTA):

- Name: —
- Contact: —
- Office: —
- Office Hours: —

Course Delivery Mode

The course delivery mode is face-to-face, and students must attend lectures. Hence, this will be no live broadcast. But, recorded lectures will be available only to a student(s) who missed that lecture due to an legitimate emergency.

Course Description

This course is intended for senior undergraduate students and graduate students interested in gaining hands-on experience applying computational techniques to solve big data analysis problems. The course topics are intentionally broad and include various data analysis skills. This way, students will become familiar with the necessary tools and understand how to perform all steps of a data analysis project.

Preferred Prerequisites

Students should demonstrate a basic understanding of algorithms and data structures, and methods to create reliable programs. Students must have a working knowledge of abstract data types and classes. Such background is needed to ensure that students can understand the sample programs provided in class and are able to write their own programs for homework assignments and class project. Students need to be familiar with Python programming language.

Course Objectives

- The course will introduce basic concepts and principal algorithms suitable for investigating basic data analysis.
- The course will help students apply and hone their data analysis skills to find meaningful patterns in the data.
- The course will give students hands-on experience in using programming languages such as Python to perform various data analysis techniques.
- In particular, the course attempts to help students answer the following essential questions:
 - How to collect, represent and store data?
 - How to pre-process data?
 - How to summarize data?
 - How to train, evaluate, and test predictive and clustering models?

Student Outcomes

- SO1 - Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions.
- SO2 - Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
- SO6 - Apply computer science theory and software development fundamentals to produce computing-based solutions.

Course Topics

Note: Some adjustments may be made during the course. Also, some additional materials might be used as well.

Topic 1: Introduction to data analysis

Topic 2: Data Representation

Topic 3: Data Collection

Topic 4: Data Storage

Topic 5: Data Quality

Topic 6: Data Preprocessing

Topic 7: Data Summarization

Topic 8: Market Basket Analysis

Topic 9: Predictive Models

Topic 10: Clustering

Topic 11: Anomaly Detection

Topic 12: Ranking (Collaborative Filtering)

Textbooks (optional)

There is no main textbook for the class. Lecture notes will be posted online. The following resources are useful.

1. Introduction to Data Mining (Pang-Ning Tan, Michael Steinbach, and Vipin Kumar), Addison Wesley, 2006.
2. Data Mining: Concepts and Techniques 3rd Edition by Jiawei Han , Micheline Kamber , Jian Pei.

Coursework

In this part, I will explain the coursework.

Assignments

The purpose of the assignments is to help you understand the concepts introduced in the lectures and to gain hands-on experience to process data. Assignments will contain both written questions and programming assignments. Below, there are some policies regarding the assignments:

- You will submit your assignment using the **Canvas System**. The assignment is due by 11:59 p.m. on the due date. **No late assignment will be accepted unless you have a legitimate excuse, e.g., an emergency.**
- You are required to do your own work. You will be responsible for ensuring that your assignment is submitted correctly. If you have multiple files to submit, you need to zip multiple files into one file before submission. If the assignment is submitted incorrectly, you can resubmit it. The system keeps track of the last file submitted for each assignment. If you do not get the confirmation information within a few minutes, you can assume the assignment has not been submitted.
- I highly encourage you to type in the answers to written questions. If you handwrite the answers, ensure they are legible and then properly scan them in a pdf file.
- At the top of the programming assignment, please include comments that consist of your name, your contact email, and the assignment name or number.
- Both undergraduate and graduate students will have the same number of assignments. Nevertheless, an assignment with the lowest score will eventually be dropped for CS5850 (undergraduates), and the average score of the rest of the assignments is considered for CS5890.

Exam

There will be a **mid-term exam** on (around end of March), covering the topics that will have been presented by then.

Project

One of the main goals of this course is to prepare you to apply data analysis tools and algorithms to an application. This is an excellent opportunity to shine your resume! **You can work in teams of up to 3 people for the Project. I highly encourage you to team up with others to perform the Project.**

Item	Description	Due Date	Grade
Proposal	You need to propose a topic for the project. (it means if you don't do it, 10% will be subtracted)	Feb 14, 202X	-10%
Progress Report	You need to write 1-2 pages on the progress of the project.	March 24, 202X	10%
Presentation	Described below	April 30, 202X	30%
Final Report	You need to write a final report in a research paper format about the project.	May 02, 202X	50%
Code & Materials	You need to submit your code and other materials.	May 02, 202X	10%

Project Presentation

Due to a rather high number of enrollments, we can't have project presentations during the lecture. Hence, you need to record your presentation (e.g., using Zoom) and upload it to the course website. This way, all students can watch the presentation offline and be informed of other projects. The uploaded presentation will be graded.

Research Paper Presentation (Bonus)

Each student can present a relevant paper during the course. The presentation will be scheduled for 15 minutes. If you decide to present a paper, you need to write a one-page summary of the paper on the day of the presentation. 60% of the grade of this assignment is for the presentation and 40% for the one-page summary. You are free to pick up the topic of the paper, but it should be within the scope of the course. You must send me the paper's title two weeks before the presentation. The deadline to turn in the one-page summary is your presentation day.

Breakdown of Points

Item	Percentage
Assignments	50%
Project	25%
Mid-term	25%
Attendance and in-class activities	-5%
Research Paper Presentation (Bonus and optional, encouraged for grad students)	<u>up to 5%</u>

Grading Scheme

Grade	Points Range
A	[93%, 100%]
A ⁻	[87%, 93%)
B ⁺	[81%, 87%)
B	[75%, 81%)
B ⁻	[70%, 75%)
C ⁺	[65%, 70%)
C	[60%, 65%)
C ⁻	[58%, 60%)
D ⁺	[55%, 58%)
D	[50%, 55%)
F	[0%, 50%)

Grading Policy and Grades

- Your score for each assignment and exam will be available through the Canvas System. You will be shown each individual score, your percent to date, and any comments. You will be notified via email whenever you receive a score or when a change is made to a score.
- If you want to dispute a score, contact David (david.guevara@usu.edu) within two days of having received the original score. If the disagreement is not resolved after communicating with the GTA, contact me by email (hamid.karimi@usu.edu). I will review your submission, talk to the GTA and then respond to you. I will not consider changes in scores if you do not contact me within one week of when the score is sent to you.