

San José State University
Department of Computer Science
CS131, Processing Big Data - Tools and Techniques, Section 2, Spring, 2023

Course and Contact Information

| | |
|------------------|---|
| Instructor: | Fabio Di Troia |
| Office Location: | MH 217 |
| Telephone: | |
| Email: | fabio.ditroia@sjsu.edu |
| Office Hours: | TT, 1:00 – 2:00pm (https://sjsu.zoom.us/j/82511129916 - PW: 720890) |
| Class Days/Time: | TuTh 3:00PM - 4:15PM |
| Classroom: | MH 422 |
| Prerequisites: | CS 46B or BIOL 123B with a grade of C- or better. Allowed Declared Majors: Computer Science BS, Data Science BS, MS Bioinformatics (MS BI). |

Course Format

Faculty Web Page and MYSJSU Messaging

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on [Canvas Learning Management System course login website](#) at <http://sjsu.instructure.com>. You are responsible for regularly checking with the messaging system through [MySJSU](http://my.sjsu.edu) at <http://my.sjsu.edu> (or other communication system as indicated by the instructor) to learn of any updates.

Course Description

In-depth study of essential tools and techniques for processing big data over the UNIX operating system and/or other operating systems. On UNIX, it includes using grep, sed, awk, join, and programming advanced shell scripts for manipulating big data.

Course Learning Outcomes (CLO)

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

- Analyze, manipulate and process large-scale data.
- Develop shell scripts for use in data-intensive applications.
- Build data analysis pipelines, automate tasks, make analyses reproducible and shareable.
- Solve big data challenges with the application of machine learning algorithms.
- Apply data science solutions to datasets from example domains, such as biology, business, finance.
- Perform big data analyses efficiently, document and reproduce analyses.

Required Texts/Readings

Textbook

None

Other Readings

None

Other technology requirements / equipment / material

None

Course Requirements and Assignments

SJSU classes are designed such that in order to be successful, it is expected that students will spend a minimum of forty-five hours for each unit of credit (normally three hours per unit per week), including preparing for class, participating in course activities, completing assignments, and so on. More details about student workload can be found in [University Policy S12-3](http://www.sjsu.edu/senate/docs/S12-3.pdf) at <http://www.sjsu.edu/senate/docs/S12-3.pdf>.

Homework, Midterm and Final exam are expected for this class. Homework is due on Canvas by class starting time on the due date. Each assigned problem requires a solution and an explanation (or work) detailing how you arrived at your solution. Cite any outside sources used to solve a problem. When grading an assignment, I may ask for additional information.

NOTE that [University policy F69-24](http://www.sjsu.edu/senate/docs/F69-24.pdf) at <http://www.sjsu.edu/senate/docs/F69-24.pdf> states that “Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading.”

CoS COVID-19 and Monkeypox Safety

Students registered for a College of Science (CoS) class with an in-person component should view the [CoS COVID-19 and Monkeypox Training](#) slides for updated CoS, SJSU, county, state and federal information and guidelines, and more information can be found on the [SJSU Health Advisories](#) website. By working together to follow these safety practices, we can keep our college safer. Failure to follow safety practice(s) outlined in the training, the SJSU Health Advisories website, or instructions from instructors, TAs or CoS Safety Staff may result in dismissal from CoS buildings, facilities or field sites. Updates will be implemented as changes occur (and posted to the same links).

Recording Zoom Classes

This course or portions of this course (i.e., lectures, discussions, student presentations) will be recorded for instructional or educational purposes. The recordings will only be shared with students enrolled in the class through Canvas. The recordings will be deleted at the end of the semester. If, however, you would prefer to remain anonymous during these recordings, then please speak with the instructor about possible accommodations (e.g., temporarily turning off identifying information from the Zoom session, including student name and picture, prior to recording). Students are not allowed to record without instructor permission. Students are prohibited from recording class activities (including class lectures, office hours, advising sessions, etc.), distributing class recordings, or posting class recordings. Materials created by the instructor for the course (syllabi, lectures and lecture notes, presentations, etc.) are copyrighted by the instructor. This university policy (S12-7) is in place to

protect the privacy of students in the course, as well as to maintain academic integrity through reducing the instances of cheating. Students who record, distribute, or post these materials will be referred to the Student Conduct and Ethical Development office. Unauthorized recording may violate university and state law. It is the responsibility of students that require special accommodations or assistive technology due to a disability to notify the instructor.

Students are required to have an electronic device (laptop, desktop or tablet) with a camera and built-in microphone. SJSU has a free equipment loan program available for students.

Students are responsible for ensuring that they have access to reliable Wi-Fi during tests. If students are unable to have reliable Wi-Fi, they must inform the instructor, as soon as possible or at the latest one week before the test date to determine an alternative. See Learn Anywhere website for current Wi-Fi options on campus.

Final Examination or Evaluation

The final test will be published on Canvas and will be submitted online.

Grading Information

- Homework, 25%
- Midterm 1, 25%
- Midterm 2, 25%
- Final Project, 25%

Note that "All students have the right, within a reasonable time, to know their academic scores, to review their grade-dependent work, and to be provided with explanations for the determination of their course grades." See [University Policy F13-1](http://www.sjsu.edu/senate/docs/F13-1.pdf) at <http://www.sjsu.edu/senate/docs/F13-1.pdf> for more details.

Determination of Grades

Semester grade will be computed as a weighted average of the scores obtained in each of the three categories listed above. No make-up tests or quizzes will be given, and no late homework (or other work) will be accepted. Also, in-class work must be completed in the section that you are enrolled in.

Nominal Grading Scale:

| Percentage | Grade |
|--------------|-------|
| 92 and above | A |
| 90 – 91 | A- |
| 88 – 89 | B+ |
| 82 – 87 | B |
| 80 – 81 | B- |
| 78 – 79 | C+ |

| | |
|--------------|----|
| 72 – 77 | C |
| 70 – 71 | C- |
| 68 – 69 | D+ |
| 62 – 67 | D |
| 60 - 61 | D- |
| 59 and below | F |

Classroom Protocol

- **Cheating** will not be tolerated.
- Student must be respectful of the instructor and other students. For example, No disruptive or annoying talking.
- Turn off cell phones
- Class begins on time
- Valid picture ID required at all times

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' [Syllabus Information web page](http://www.sjsu.edu/gup/syllabusinfo/) at <http://www.sjsu.edu/gup/syllabusinfo/>". Make sure to review these policies and resources.

CS131 / Processing Big Data - Tools and Techniques, Spring 2023

Course Schedule

The schedule is subject to change with fair notice communicated via Canvas course page

Course Schedule

| Week | Date | Topics, Readings, Assignments, Deadlines |
|------|-------|--|
| 1 | 01/26 | Introduction |
| 2 | 01/31 | Introduction |
| 2 | 02/3 | Machine Learning fundamentals |
| 3 | 02/7 | Machine Learning fundamentals |
| 3 | 02/9 | Theory: Machine Learning for Big Data: Supervised approach |
| 4 | 02/14 | Lab: Machine Learning for Big Data: Supervised approach |
| 4 | 02/16 | Theory: Machine Learning for Big Data: Supervised approach |
| 5 | 02/21 | Lab: Machine Learning for Big Data: Supervised approach |
| 5 | 02/23 | Theory: Machine Learning for Big Data: Supervised approach |
| 6 | 02/28 | Lab: Machine Learning for Big Data: Supervised approach |
| 6 | 03/2 | Lab: Data Preprocessing |
| 7 | 03/7 | Midterm 1 – Week |
| 7 | 03/9 | Midterm 1 – Week |
| 8 | 03/14 | Theory: Machine Learning for Big Data: Unsupervised approach |
| 8 | 03/16 | Lab: Machine Learning for Big Data: Unsupervised approach |
| 9 | 03/21 | High-class Imbalance in Big Data: Tools and techniques |
| 9 | 03/23 | Theory: Machine Learning for Big Data: Unsupervised approach |
| 10 | 03/28 | SPRING BREAK |
| 10 | 03/30 | SPRING BREAK |
| 11 | 04/4 | High-class Imbalance in Big Data: Tools and techniques |
| 11 | 04/6 | High-class Imbalance in Big Data: Tools and techniques |
| 12 | 04/11 | High-class Imbalance in Big Data: Tools and techniques |
| 12 | 04/13 | Midterm 2 - Week |
| 13 | 04/18 | Midterm 2 - Week |
| 13 | 04/20 | TBD |
| 14 | 04/25 | TBD |

| Week | Date | Topics, Readings, Assignments, Deadlines |
|-------------|-------------|---|
| 14 | 04/27 | TBD |
| 15 | 05/2 | TBD |
| 15 | 05/4 | Project Presentation |
| 16 | 05/9 | Project Presentation |
| 16 | 05/11 | Project Presentation |
| Final Exam | 05/19 | 2:45 – 5:00 PM |