Idaho State University | Informatics and Computer Science

Data Science and Applied Machine Learning (3 credits)

Spring Semester 2023

CS 4499/5599

Monday and Wednesday

3:00 pm – 4:15 pm MT

ISU: CHE 310 (IF) / LIB 13 (Pocatello)/Zoom

Instructor:	Pepo Mena
Office:	CAES 217
Email:	menapedr@isu.edu/pepomena@isu.edu
T/A:	Kaden Marchetti
Email:	kadenmarchetti@isu.edu
Tutoring Hours:	By email and appointment

Syllabus

COURSE DESCRIPTION

This course covers intermediate subjects in data science and machine learning using python data science stack, including pandas, NumPy, and scikit-learn. The following machine learning models will be covered: multiple linear regression, logistic regression, k-nearest-neighbors, naive Bayes, support vector machines, decision trees and random forests, principal component analysis, and k-means clustering. Exploratory data analysis, preprocessing techniques, feature engineering, dimensionality reduction, ensemble methods, and evaluation of models will also be covered.
A review of python, NumPy, and pandas will be given at the beginning of class.

COURSE PREREQUISITES

CS 1181: Introduction to Programming (Experience with Python is strongly recommended)
MATH 1170: Calculus 1

TEXTBOOK

Required: None

Optional: Hand-On Machine Learning with Scikit-Learn , Keras & TensorFlow by Aurelien Geron

Please note that a Google Collaboratory Pro subscription may be useful, but will not be required

COURSE OBJECTIVES

Data Importing & Exploration

Students will learn how to import datasets into a Python environment from a variety of sources including from files of different formats and web resources such as GitHub. Students will also learn how to manipulate, query, and perform simple operations with said data using tools such as pandas.

Data Preparation and Preprocessing

Students will learn the importance of data preparation in the development of machine learning models. We will cover a variety of statistical methods found in the scikit-learn. These include normalization, robust scaling, and Principal Component Analysis.

Machine Learning Techniques

This class will focus on the scikit-learn Python library. We will cover both regression and classification models and how to evaluate the quality of the model trained. Techniques covered will include: linear regression, decision trees/random forest, k-nearest neighbors, k-means, among others. There will also be a discussion on ethics in data science.

ASSIGNMENTS

Homework assignments will be given regularly. A midterm project and final project will also be assigned. Homework and projects give you as the student the opportunity to apply and master the material covered in class and is an essential aspect of the course. To help prepare the graduate students for research they will create their own final project. They must submit a preliminary proposal, written report, and presentation. More information will be given later. The university also requires that Graduate students perform additional work above the undergraduate level, as such graduate students should expect to have some additional work on assignments and projects. Although homework is expected in a timely manner, late homework will be accepted late. Late homework will forfeit 10% of the grade if overdue less than one week, and 20% of the grade if overdue more than one week. Submissions later than two weeks must be approved by the instructor.

TUTORS

A tutor is available to help students gain any additionally needed understanding of course material to be able to complete assignments.

ABSENCES

Attendance is highly encouraged. If for some reason you are absent from a class, it is your responsibility to contact your fellow students for what you missed in class. This includes homework assignments that were assigned, and additional content discussed.

EXPECTATIONS ON PROFESSIONALISM AND ATTITUDE

Be prepared for class. Read the assigned readings carefully before class, and spend some time thinking about what you've read. Write down questions about the reading as they occur to you. Get to class on time.
Ask questions in class.
Stay alert and attentive in class.
Participate in class discussions in a courteous and positive manner.
Do not carry on "sidebar" conversations during instructor, student or guest presentations.
Laptops can be used to take notes, review notes or documents relevant to that day's class, or to quickly search for information to answer a question that has come up during discussion. If you are discovered to be reading email, participating in on-line chats, surfing the web or other non-class related activities using laptops or other accouterments during class, you will be warned and upon repeated occurrences requested to leave the class.
Spend enough time on assignments to ensure that you are actually learning the material and are able to turn in work of high quality. Assignments are an opportunity to learn, not something to just "get done."
Turn in assignments on time.

ACCOMMODATION

Students with disabilities are encouraged to discuss their needs with the instructor, preferably during the first week of class. All reasonable accommodations will be made to see that disabilities do not restrict a student's opportunity to learn. Students are also highly encouraged to call the ADA and Disabilities Resource Center at 282-3599.

ACADEMIC INTEGRITY

Academic integrity is expected at Idaho State University and the College of Business. All forms of academic dishonesty, including cheating and plagiarism, are strictly prohibited, the penalties for which range up to permanent expulsion from the university with "Expulsion for Academic Dishonesty" noted on the student's transcript. If you are unclear as to what constitutes academic dishonesty, you can get a copy of the College of Business Policy on Academic Integrity from the College of Business office in BA 202, or from the College of Business website at www.cob.isu.edu, or refer to the ISU Faculty/Staff Handbook policy on academic dishonesty at:
http://www.isu.edu/policy/fs-handbook/part6/6_9/6_9a.html or
http://www.isu.edu/library/research/ait/aitsitemap.html
Academic dishonesty includes, but is not limited to 1) cheating on examination, 2) plagiarism, or 3) collusion.

Definitions:
Cheatingon an examination includes:

- Copying from another's paper, any means of communication with another during an examination, giving aid to or receiving aid from another during an examination;
- Using any material during an examination that is unauthorized by the proctor;
- Taking or attempting to take an examination for another student or allowing another student to take or attempt to take an examination for oneself.
- Using, obtaining, or attempting to obtain by any means the whole or any part of an un-administered examination.
- Talking to anyone other than the professor during an exam.

Plagiarism is the unacknowledged incorporation of another's work into work which the student offers for credit.

Collusion is the unauthorized collaboration of another in preparing work that a student offers for credit.

Other types of academic dishonesty include:

- Using other student's content from the labs or students' disk, etc.
- Performing any act designed to give unfair advantage to a student or the attempt to commit such acts is considered cheating.

Notes:

- The use of the source code of another person's program, even temporarily, is considered plagiarism.
- Copying material from a source without attributing (citing) the source.
- Allowing another person to use your source code, even temporarily, is considered collusion.

In this class, the specific exceptions given below are not considered scholastically dishonest acts:

- Discussion of the algorithm and general programming techniques used to solve a problem.
- Giving and receiving aid in debugging
- Discussion and comparison of program output (output only not code)

If the instructor finds an assignment that has been submitted as your own work and is not, it will be recorded as a 0 for all parties involved. If this is repeated the grade will once again be recorded as a 0 for all involved and a report will be filed with the university. If you are found cheating on an exam, your exam will be taken and your result for the exam will be recorded as a 0. All instances of cheating on exams will be referred to the university. The university may take other action.

SUCCESS

Success in this course depends heavily on your personal health and well-being. Recognize that stress is an expected part of the college experience, and it often can be compounded by unexpected setbacks or life changes outside the classroom. I encourage you to reframe challenges as an unavoidable pathway to success. Reflect on your role in taking care of yourself throughout the term, before the demands of exams and projects reach their peak. Please feel free to reach out to me about any difficulty you may be having that may impact your performance in this course. If you are experiencing stress in other areas of your campus life, I am happy to help you get in contact with other resources on campus that stand ready to assist you. In addition to your academic advisor, I strongly encourage you to contact the many other support services on campus that are available.

SERVICES

ISU Counseling and Testing Services (CATS) would like to remind all students who are enrolled in the current semester (part-time or full-time) they are eligible for free, confidential counseling services. CATS offers individual, group, and couples counseling, as well as Biofeedback Training. We also offers crisis intervention services Monday through Friday from 8-5.

Locations:

Pocatello: Graveley Hall, 3rd floor of the south side. To schedule an appointment call 208-282-2130 or just stop by. Walk-ins are welcome Monday-Friday from 8am - 4pm.

Idaho Falls: Bennion Student Union, 2nd floor, Room 223. Please call 208-282-7750 to schedule an appointment.

Meridian: Counseling services are available to Meridian students through the ISU Counseling Department. Please call 208-373-1719 to schedule an appointment.

COMFORTABLE LEARNING ENVIRONMENT

We are all committed to maintaining an inoffensive, non-threatening learning environment for every student. Class members (including the instructor) are thus to treat each other politely—both in word and deed. Offensive humor and aggressive personal advances are specifically forbidden. If you feel uncomfortable with a personal interaction in class, see your instructor for help in solving the problem.

GRADING

Course grading will be determined using the following breakdown:

Topic	Value
Homework	60%
Midterm/ Mid-project	20%
Final	20%

Tentatively, grades will be earned based on a straight scale grading policy. The instructor reserves the right to change the grading scale, and assignment weighing. Such changes would be:

i) based on professional judgment
ii) applied across the board to all students
iii) in favor of the students

TENTATIVE COURSE SCHEDULE

Date	Topic
1/10-1/12	DS and ML Overview, and Python NumPy
1/17 – 1/19	Pandas
1/24 – 1/26	Data Visualization
1/31 – 2/2	Statistical Inference
2/7 – 2/9	Regression
2/14 – 2/16	Decision Trees
2/21 – 2/23	Ensemble Methods and Random Forests
2/28 – 3/2	Logistic Regression
3/7 – 3/09	k-Nearest-Neighbor
3/14– 3/16	Support Vector Machines
3/21 – 3/23	No Classes “Spring Break”
3/28 – 3/30	Naive Bayes
4/4 – 4/6	Clustering
4/11 – 4/13	Dimensionality Reduction
4/18– 4/20	Hyper Parameter Tunning
4/25-4/27	Auto Machine Learning
5/02-5/04	Final Exam