



CAL POLY
College of Engineering

Spring 2022 IME 372: Applications of Enterprise Analytics

INSTRUCTOR

Puneet Agarwal, Ph.D.

Office: Room 232, Bldg. 192

E-mail: pagarw05@calpoly.edu (Please put "IME 372" on the subject of all your emails).

Phone: 805-756-7467 (Office)

Office Hours: Mondays and Wednesdays, 3:10 PM - 5:00 PM, In-person, 192-232

Zoom Link: <https://calpoly.zoom.us/my/pagarw05>

COURSE DESCRIPTION

Applications of Big Data Analytics to facilitate enterprise decision-making. Clustering, classification, and prediction. Applications of regression-based prediction and decision making. Over-fitting and regularization. Supervised and unsupervised learning. Machine learning, neural networks, and Bayesian analysis. 3 lectures, 1 lab.

Class Meetings:

- ★ *Lecture* - Tue/Thu, 1:40 PM to 3:00 PM, Room 221, Bldg. 192
- ★ *Lab* - Fri, 8:10 AM to 11:00 AM, Room 240, Bldg. 192

Prerequisites: IME 212, IME 312, IME 326, MATH 244.

What You Can Get out of This Course: Over the last decade, organizations across industries have come to rely on the 2.5 quintillion bytes of data humans generate daily to understand their customers better, identify patterns in behavior, and make more effective and strategic decisions. As data becomes more prevalent across many different areas of importance in engineering, policy analysis, and management, analytics is becoming an increasingly important topic. This course will focus on building competencies that enable students to leverage data analytics and predictive modeling to optimize business performance at a variety of levels in a wide range of industries such as health care, manufacturing, and cybersecurity. Through several case studies, students will learn how to identify situations where data analytics and predictive modeling could be used, as well as learn tools and acquire skills for data manipulation, extraction, visualization, analytical modeling, and exploiting patterns found in historical data for identifying risks and opportunities. Since predictive models are now being leveraged to make important decisions, it becomes paramount to keep models from making unfair predictions. From human bias to dataset awareness, students will also explore many aspects of building more ethical models.

Learning Objectives: After successfully completing this course, you should be able to:

- ★ Identify enterprise opportunities to improve through predictive model building, and specify business value-added proposition.
- ★ Select the appropriate data mining tasks (clustering, classification, or prediction) to address the identified enterprise opportunity.

- ★ Identify and utilize the appropriate machine learning algorithms to solve enterprise analytic problems.
- ★ Evaluate the capabilities and weaknesses of different machine learning algorithms, and choose the one that best fits the case at hand.
- ★ Recognize and explain the difference between supervised and unsupervised learning algorithms and employ an appropriate algorithm for the case in hand.
- ★ Conduct both traditional and modern regression analyses as predictive model building.
- ★ Conduct appropriate model fitting and analyze the generalizability of the results.
- ★ Assess and make sense of predictive model analysis/results and correlate the results to the enterprise original business proposition.

Course Website: The course materials including lecture notes, the syllabus, and assignments will be accessible through [Cal Poly Canvas](#). Students should regularly check the course website for announcements and uploaded materials.

Computing: In this course, we will make extensive use of [Google Colab](#). Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute Python code through the browser (Chrome, Firefox, and Safari), and is especially well suited to data analysis, machine learning, and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs. As a programmer, you can perform the following using Google Colab:

- ★ Write and execute code in Python
- ★ Document your code that supports mathematical equations
- ★ Create/Upload/Share notebooks
- ★ Import/Save notebooks from/to Google Drive
- ★ Import external datasets, e.g., from Kaggle
- ★ Free Cloud service with free GPU

REFERENCES

There is no textbook for this course. The course notes will be distributed via Cal Poly Canvas. Recommended reading sources are as follows:

- ★ *Python Data Science Handbook: Essential Tools for Working with Data* by Jake VanderPlas, Publisher(s): O’Reilly Media, Inc., ISBN: 9781491912058.
- ★ *Python for Data Analysis*, 2nd Edition by Wes McKinney, Publisher(s): O’Reilly Media, Inc., ISBN: 9781491957660.
- ★ *Introduction to Machine Learning with Python* by Andreas C. Müller, Sarah Guido, Publisher(s): O’Reilly Media, Inc., ISBN: 9781449369415.
- ★ *Machine Learning with Python Cookbook* by Chris Albon, Publisher(s): O’Reilly Media, Inc., ISBN: 9781491989388.
- ★ *Fundamentals of Machine Learning for Predictive Data Analytics*, 2nd Edition: Algorithms, Worked Examples, and Case Studies by John D Kelleher, Brian Mac Namee, Aoife D’Arcy, Publisher(s): The MIT Press, ISBN: 9780262044691.

COURSE GRADING

The following scale represents the minimum course average needed for you to be guaranteed each letter grade: A (95); A- (90); B+ (87); B (85); B- (80); C+ (77); C (75); C- (70); D+ (67); D (65); and D- (60). I reserve the right to lower these cutoffs (i.e., give higher grades than indicated) under some circumstances. However, I will not raise the cutoffs. I also follow the University's policy on Explanation of Grades, see <https://catalog.calpoly.edu/academicstandardsandpolicies/grading/>.

ASSESSMENTS

Student learning will be assessed using the following methods: homework assignments, lab exercises, a final exam, and a course project. Your grade will be calculated using the following distribution:

Item	Percentage
Homework assignments	20%
Lab exercises	10%
Final exam	25%
Project proposal	10%
Project final presentation	15%
Project final report	20%

Homework: Homework assignments are designed to reinforce the material discussed in the lectures and lab sessions. You are encouraged to join with other students in discussing the course, including homework. This is especially useful if you have first tried to solve the problem on your own, and focus on understanding the reasons for any differences between your answer and someone else's approach, rather than just copying the answer that someone else obtained. Note that each student must submit an individual copy for homework assignments unless stated otherwise. The due date and grading policy for each homework assignment will be clearly communicated. Late homework will not be accepted except in extenuating circumstances (e.g. family emergency, illness, etc.), with official documents and prior instructor permission. For programming problems, you will need to turn in code in Jupyter Notebooks. For written assignments, you are strongly recommended to turn in a computer-generated (e.g., \LaTeX , MS Word) homework if possible.

Lab Exercises: In each lab session, students will work on programming exercises and applied problems on predictive analytics that are designed to reinforce the concepts discussed in lectures. Students are expected to complete and submit these assignments by the end of each session.

Final Exam: There will be one comprehensive final exam from *1:10 PM to 4:00 PM on Tuesday, June 7, 2022*. The exam will have multiple-choice, essay, and math questions. Makeup exam will be administered only under extenuating circumstances with official documents, provided that I am notified in advance.

Project: Each group (1-3 students) works on a project. Project is an essential part for this class, which consists of three parts: proposal, final presentation, and report.

- ★ **Proposal:** 2-3 pages, single spacing with 12pt fonts; including title, team members, abstract, description of the enterprise problem, summary of the data to be used, project plan with timelines, coordination plan between group members, and references.
- ★ **Final Presentation:** ~ 15-18 minutes, evaluated by both the instructor and peers.
- ★ **Report:** 10-15 pages, single spacing with 12pt fonts; including title, team members, abstract, description of the enterprise problem, summary of the data (attributes description, statistical summaries, visualizations), data cleaning and preprocessing, methodology (supervised/unsupervised learning), model selection, performance evaluation, results, assumptions, ethical considerations, conclusions, a statement of individual contributions of team members, and references.

Students may choose their own project or consult the instructor for project suggestions. The project may be from your internships, senior projects, or master's thesis/project. The delivery of the project does not have to be solely for this class, and it can be used for other classes and purposes. Each group is required to discuss with and get a permission from the instructor, regarding the detailed project topics and plans, in email, office hours or by appointments.

Item #	Description	Submission Date
1	Topic Selection and Team Formation	April 15, 2022 (Week 3)
2	Project Proposal	April 22, 2022 (Week 4)
3	Project Final Presentation	May 27, 2022 (Week 9)
4	Project Final Written Report	June 3, 2022 (Week 10)

OFFICE HOURS

To be fair to all of the students and to the colleagues I work with, please cooperate in respecting my office hours. However, individual appointments can be made, if the posted office hours are insufficient or inconvenient. You may email me at any time to discuss your questions, or to set up an appointment. I will respond as soon as possible, but you should be aware that I do not always check my email on the evenings and weekends, or when I am traveling.

UNIVERSITY POLICY ON ACADEMIC INTEGRITY

Cal Poly will not tolerate academic cheating or plagiarism in any form. Academic dishonesty is addressed both as an academic issue and as a disciplinary incident under the CSU Standards for Student Conduct. Cases of class cheating or plagiarism shall be handled by faculty members under established procedures that include written notice to the student of the incident and the consequent grade. This response is, by its nature, limited to the particular class incident. The faculty member shall then submit an online report to the Office of Student Rights & Responsibilities. This office will consider the reported incident and the academic response in the broader context of the student's overall conduct. Please thoroughly familiarize yourself with the [Office of Students' Rights and Responsibilities page on Academic Integrity](#), as well as the information found on [Academic Programs and Planning website regarding cheating and plagiarism](#).

CRITICAL CAMPUS RESOURCES

Sexual Harassment and Violence: Cal Poly is committed to fostering a safe, productive learning environment for all students. Title IX and CSU policy prohibit discrimination on the basis of sex, which includes sexual harassment, domestic and dating violence, sexual assault, and stalking. We understand that sexual violence can undermine students' academic success and we encourage students who have experienced some form of sexual misconduct to access appropriate resources so they can get the support they need and deserve.

As an instructor, I have a mandatory reporting responsibility as a part of my role. It is my goal that you feel comfortable to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep information you share private to the greatest extent possible. However, I am required to report information I receive regarding sexual misconduct or information about a crime that may have occurred during your time at Cal Poly. There are several on campus resources available to support survivors. Please see the list below:

Students can *report incidents* of alleged sexual misconduct to either or both of the following resources:

- * Office of Equal Opportunity (Title IX) — equalopportunity@calpoly.edu — 805-756-6770
- * Cal Poly Police Department — police@calpoly.edu — 805-756-2281

Students can access *confidential support* from two separate resources on campus:

- ★ Safer — safer@calpoly.edu — 805-756-2282
- ★ Counseling Services — counseling@calpoly.edu — 805-756-2511

A 24-hour confidential hotline for survivors is also available through the [local San Luis Obispo non-profit RISE](#) at 855-886-7473.

COVID-19 Compliance, Classroom, and Campus Safety: Cal Poly is committed to protecting the health and safety of the campus community. Taking preventative steps, as well as monitoring your health and staying home if you are feeling unwell, will help protect the entire Cal Poly community.

By participating in this course, you agree to abide by all campus safety protocols. Please note that safety protocols may change throughout the quarter. You must follow all protocols as outlined in the most recent [campus updates](#).

Accessibility Resources: It is University policy to provide, on a flexible and individualized basis, reasonable accommodations to students who have disabilities that may affect their ability to participate in course activities or to meet course requirements. If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and the Disability Resource Center, Building 124, Room 119, at 805-756-1395, as early as possible in the term. [Apply for Disability Services](#) or email drc@calpoly.edu.

Inclusion and Classroom Climate: I strive to make this classroom a place where you will be treated with respect, and I welcome individuals of all ages, backgrounds, beliefs, races, ethnicities, social classes, genders, gender identities, gender expressions, national origins, documentation statuses, religious affiliations, sexual orientations, abilities – and other visible and nonvisible differences. All members of this class are expected to contribute to a respectful and inclusive environment for every other member of the class. This does not mean we cannot disagree or have different ideas. It does mean we try to consider perspectives other than our own, though they may differ from our own beliefs/experiences. If you experience disrespect or discrimination in this class, please report your experiences to me.

Well-being and Mental Health Support: Recent nationwide surveys of college students consistently find that stress, sleep problems, anxiety, depression, interpersonal concerns, death of a significant other and alcohol use are among the top ten health impediments to academic performance. Students experiencing personal problems or situational crises are encouraged to contact Cal Poly's Counseling Services (805-756-2511) for assistance, support and advocacy. This service is free and confidential. Please use the following link to get updated information related to health and well-being: <https://coronavirus.calpoly.edu/>.

COURSE POLICIES

To create and preserve a classroom atmosphere that optimizes teaching and learning, all students share the responsibility of creating a positive learning environment. Students are expected to conduct themselves in a manner that does not disrupt teaching or learning, and they are expected to follow these standards:

- ★ Because random arrivals and exits are disrespectful and distracting, please plan to arrive to class on time and to stay for the entire class period. If circumstances dictate that you must be late or you know that you will need to leave early, please take a seat close to the door so you do not distract others during class time.
- ★ Students are expected to refrain from use of cell phones or other electronic devices unless they are clearly linked to class purposes (e.g., note-taking).

- ★ Talking and other disruptive behaviors are not permitted while lecture is taking place. Please pay attention, and do not hesitate to ask me any questions.
- ★ Classroom discussion should be civilized and respectful to everyone and relevant to the topic we are discussing. Classroom discussion is meant to allow us to hear a variety of viewpoints. This can only happen if we respect each other and our differences.
- ★ Students are expected to use professional style throughout the class and in all communications, including emails to the instructor. This includes the use of salutations and closings (including clear identification of the author) and correct grammar.

TOPICS TO BE COVERED

1. Data Processing and Analysis using NumPy and Pandas
2. Data Visualization and Geographic Plotting using matplotlib and folium
3. Introduction to Predictive Analytics using Machine Learning
 - ★ Types of Machine Learning Systems
 - ★ Working Principle of Machine Learning
 - ★ Challenges of Machine Learning
4. Model Selection and Performance Evaluation
5. Supervised Learning Models
 - ★ Multiple Linear Regression
 - ★ Lasso and Ridge Regression
 - ★ Tree-based Models
 - ★ Ensemble Learning using Bagging and Boosting
 - ★ Support Vector Machines
 - ★ Artificial Neural Networks
6. Unsupervised Learning Models
 - ★ Clustering
 - ★ Principal Component Analysis
7. Data Ethics and Model Biases