



BDS4440 Advanced Programming for AI and Big Data

Course number/section: BDS4440 Instructor Name: Joe Ganser
Course name: Advanced Programming for AI Big Data and Cloud Computing
Prerequisite: BDS 2240 Instructor Email:
Credit Hours: 3 Credits Office Hours:
Classroom Number:
Syllabus Author: Joe Ganser

COURSE DESCRIPTION

This course is designed to provide students with advanced machine learning and Python programming skills to learn today's most compelling leading-edge computing technologies including AI, big data and cloud case studies on natural language processing, machine learning, deep learning, computer vision, Hadoop, Spark and the internet of things.

COURSE LEARNING OBJECTIVES

At the successful conclusion of this course, students will be able to:

1. Identify broad opportunities for automation with machine learning and big data
2. Apply advanced Python programming and computer science thinking skills
3. Practice hands-on library-focused applied approach using the interactive iPython interpreter and code
4. Use modern natural language processing tools, formulations and problems
5. Use deep learning tools, formulations and problems
6. Apply Big data and internet of things concepts and tools
7. Describe current trends in the field and opportunity that they bring

REQUIRED RESOURCES

To improve student learning and provide students with a more enjoyable academic experience, a customized eText and/or digital learning resources are provided for this course. These customized learning resources are preloaded into this course and are accessible through the course Canvas site.

For assistance with issues related to customized eText access, contact:
helpdesk@BerkeleyCollege.edu or call 973-278-5400 ext. 1540.

Texts

1. *Required:* Intro to Python for Computer Science and Data Science By Paul & Harvey Dietel 2020 ISBN10: 0-13-540467-3
2. *Optional:* Introduction to Data Mining, 2nd edition 2018 By PangNing Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar ISBN ISBN-10: 0133128903

GRADING POLICY

All assignments are submitted on Canvas by the due date/time required. All grading is done by the rubrics. Please check your Canvas grade book regularly and review grading feedback. Feel free to contact me during my office hours if you have any question or concern about your grades.

Grades in this course are based on a weighted average, as follows:

%	Component	Learning Objectives
10	Class Discussion	#1, #2, #7
50	Homeworks	#1-#7
40	Final Project	

- Homeworks will be submitted via jupyter notebooks
- Must attempt every problem! Do your best, partial credit is given
- Must complete every homework.
- Unsubmitted homeworks WON'T be dropped!

Class Discussion

Each week there will be discussion questions. Students should answer the question and participate in feedback of other student's responses.

40% Final Project

No tests - just homeworks and projects! Use any resources needed, just don't plagiarize. Always cite sources properly.

FINAL PROJECT

Final Project will be the presentation of a data science project in a web based format.

It must include a presentation of analysis on data. I will give reference links to data sources.

It will include two major components:

1. Business person presentation (for a non technical audience)
 - a. Present in both Jupyter Notebook and Google Slides format
2. Code based analysis (for a computer science audience)
 - a. Present in Python file format, with text describing the role of code.

Main objectives of the project:

- Use data science and or machine learning techniques to solve a real world business/scientific problem that adds value to society.
- Be able to explain it to a non-scientist.
- Be able to justify your scientific reasoning.
- Present it in easy to read python code that will produce the same results when running on another computer.

Project Criteria:

- Business Presentation is easy to understand, referenced, logical and insightful
- Scientific Presentation is logical, follows known scientific principles, and solves a real world problem.
- Code is well written, documented, justified in its use and easy to understand
- Project relates to data science techniques we used in class

Options of the project

- A. Use a common machine learning repository with known data sets to come to scientific conclusions
- B. Compete in a kaggle/driven data or other data science competition
- C. Work in teams of 2-4 people, with defined roles, or work independently.

See my own blog for reference example: <https://JoeGanser.github.io>. To pass the course, you must pass & complete the final project.

Final Project Grading Rubric			
Proficient (90-100 %)	Competent 70-80%	Novice 60-70%	Below expectations Below 60%
Contextual information was relevant and clearly explained. Clear evidence of a thoughtful approach to the project. Every analytical technique is given an accurate justifiable reasoning for its use. Sources cited, code well written	Contextual information was relevant and sufficiently explained. Some evidence of a thoughtful approach to the project, but not all techniques are justified using the scientific method. Sources cited, and code is decent but not at professional level.	Contextual information was brief and/or vague. Limited evidence of a thoughtful project approach	Contextual information lacks relevance and quality No evidence of a thoughtful approach to the project.

GRADING SCALE

100-90%	A		79-75%	C+
89-85%	B+		74-70%	C
84-80%	B		69-60%	D
Below 60 = F				

Class will meet 2x a week, for an hour and a half each time via ZOOM/in person.

Office hours by appointment

Syllabus

Week	Topic	Text Source	Assessments
1	Files and Exception Text files, JSONs, pickle files, try except finally clauses, raising errors, CSV files	Ch 9	Ch9 Problems 1-6
2	Object Oriented Programming Classes, class attributes, properties of data access Simulating private attributes, Case studies, Inheritance, inheritance hierarchy, polymorphisms, docstrings, namespaces, scope, simple regression	Ch 10-10.17	Ch10 Problems 2,3,4,6,9 Ch9 Problem 7

3	Recursion, Searching, Sorting and Big O Factorials, recursion, iteration, searching, sorting, linear search, algorithm efficiency	Ch 11-11.8	Ch11 Problems 1-6 Ch10 problem 12 Ch9 problem 16
4	Recursion, Searching, Sorting and Big O Binary Search, sorting, selection, insertion, merging, visualizing Natural Language Processing (NLP) Text blob, Visualizing Word frequencies with bar charts and word clouds	Ch 11.9-11.16 Ch 12-12.3	Ch11 Problem 8,13,14 Ch12 Problems 1,2,3 Ch9 Problem 16
5	Natural Language Processing (NLP) Text blob, Visualizing Word frequencies with bar charts and word clouds	Ch 12.3-12.14	Ch12 Problems 4-12 Ch 10 Problem 18
6	Data Mining Twitter Twitter API, Creating an App, Tweepy, Getting twitter data, Searching tweets	Ch 13-13.10	Ch 13 Problems 1-10 Ch 11 Problem 18
7	Data Mining Twitter Spotting twitter trends, cleaning tweet data, streaming twitter APIs, sentiment analysis, geocoding and mapping, storing tweets, time series	Ch 13.11-13.18	Ch 13 Problems 11-16 Ch 12 Problem 17
8	Machine Learning: Classification, Regression and Clustering Types of ML, Classification case studies, Time series and regression	Ch 15.1-15.4	Ch 15 Problems 3,4,6,7,8,11,12 Ch10 Problem 11
9	Machine Learning: Classification, Regression and Clustering Multiple Regression, Dimensionality Reduction, Kmeans	Ch15.5-15.8	Ch 15 Problems 1,2,5,15,17,18 Ch 11 Problem 19
10	Deep Learning Keras, Neural Networks, Tensors, CNNs	Ch16.1-16.6	Ch16 Problems 1-6 Ch15 Problem 19

11	Deep Learning Visualizing Neural Networks, RNNs, Tuning Neural Networks, Pretrained Models and Reinforcement Learning	Ch16.7-16.13	Ch16 Problems 7-13 Ch15 Problem 20
12	Big Data SQL, NoSQL and MongoDB	Ch 17.1-17.4	Ch17 Problems 1-7 Ch13 Problem 18
13	Big Data Hadoop, Spark, IoT	Ch 17.5-17.9	Ch17 Problems 8-14 Ch 16 Problem 15,16
14	Association Rules: Frequent Itemset Generation, Rule Generation, Compact Representation of frequent itemsets, FP-Growth Algorithm, Evaluation of Association Patterns	Second Text Ch 6	Ch6 Second Text Problems 1-6 ==== Ch15 (first text) Problem 29 Ch11 Problem 20
15	Unsupervised Learning: K means, Agglomerative Clustering and DB Scan	Second Text: Ch 8	
	Final Project Presentation		

BIBLIOGRAPHY OF REFERENCE TEXTS

1. *Required:* Intro to Python for Computer Science and Data Science By Paul & Harvey Dietel 2020 ISBN10: 0-13-540467-3
2. *Optional:* Introduction to Data Mining, 2nd edition 2018 By PangNing Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar ISBN ISBN-10: 0133128903
3. Hands on Machine Learning with Scikit-Learn & Tensor Flow, March 2017, Aurelien Geron, ISBN: 9781491962299
4. Learning Python, Mark Lutz, 5th edition, 2013, ISBN-10: 1449355730

5. The Elements of Statistical Learning, 2nd Edition, Hastie/Tibshirani/Friedman, 2009, ISBN-10: 0387848576
6. Natural Language Processing with Python, Bird, Klein, Loper, 2009, ISBN-10: 0596516495
7. Introduction to Machine Learning with Python by Andreas C Muller, 2017, ISBN-10: 1449369413
8. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, by Wes McKinney, ISBN-10: 1491957662**
9. <https://machinelearningmastery.com>

Compliance with the Americans with Disabilities Act (ADA)

Berkeley College complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990, as amended by the ADA Amendments Act of 2008. Any student who seeks a reasonable accommodation of a disability with respect to an academic matter should obtain a Berkeley College *Request for Accommodation of Disability Form*, as soon as the need becomes apparent, from one of the following ADA Coordinators: 1) NJ Campuses and OL Campus: Sandra Coppola – (973) 278-5400, ext. 1320, sec@BerkeleyCollege.edu; 2) NY Campuses: Diane Georges - (212) 986-4343, ext. 4216 diane-georges@berkeleycollege.edu or **Disability Services directly at DisabilityServices@berkeleycollege.edu**. The student should specify on this form the accommodation sought, as well as the reason for and duration of the need, and attach appropriate supporting documentation when submitting this signed form to one of the ADA Coordinators. The **eligibility** for an academic accommodation will be determined based upon the level of disability, its impact on learning, and the College's ability to provide the accommodation without incurring undue burden or fundamentally altering its programs, facilities, policies, or activities. Specific details of the disability will remain confidential between the student and the ADA Coordinator/Director of Disability Services, unless the student chooses to disclose, or there is a legitimate academic need for disclosure, on a case-by-case basis."

CREDIT HOUR ASSIGNMENT POLICY

Course work performed outside of the classroom (such as reading, studying, writing papers, doing projects or receiving tutoring) is critical to academic success. While the time requirements for individual students may vary somewhat, a general rule of thumb is that students should spend two hours outside the classroom for every hour required in it. For more information about this policy, review the full credit hour assignment policy in the Berkeley College Undergraduate Catalog.

ACADEMIC INTEGRITY

Berkeley College is committed to providing an educational experience designed to develop professional competencies, including habits of personal and professional integrity. The College expects all members of its community – students, faculty, and staff – to act honestly in all situations. Actions of Academic Dishonesty will not be tolerated. "Academic dishonesty (is any) form of cheating and plagiarism which result in students giving or receiving unauthorized

assistance in an academic exercise or receiving credit for work which is not their own." (Kibler et. al., 1988, Academic integrity and student development: Legal issues and policy perspectives, Ashville, NC: College Administration Publications, Inc., p. 1.) All students are expected to agree to a pledge of honesty concerning their academic work, and faculty are expected to maintain the standards of that pledge.

TURNITIN AS A LEARNING TOOL

Turnitin helps prevent plagiarism by providing both the student and the professor a feedback report that compares any student work submitted through the software with a comprehensive database of books, journals, websites and papers written by other students. Some of the writing assignments in this course will use Turnitin software to help students improve their skill at paraphrasing statements contained in research on a topic and to help increase awareness of the proper use of citation when a student writes a paper using ideas or statements taken from a research source. For any essay-based assignment submissions, including but not limited to homework, papers, and projects, students will be expected to submit that assignment through Turnitin in Canvas, following the submission guidelines given with the assignment instructions. Prior to submitting a final draft of an assignment, students will have the opportunity to submit several drafts of that assignment to Canvas in order to get sufficient feedback from Canvas reports to help minimize the risk of plagiarism. If the assignment continues to have evidence of plagiarism in the final draft of the assignment, the professor will file a report to the Department Chair documenting the use of the paper as an action of academic dishonesty. If a student fails to submit an assignment to Turnitin, the professor will assign a grade of zero for that assignment. By submitting a paper to Turnitin, that paper will become source material included in the Turnitin database.

PROGRESS REPORTS

At the end of Weeks 4, 7 and 11, students whose course performance needs improvement will be notified through their Berkeley email. These progress reports may include specific strategies for course improvement and your options if you are at risk for course failure.

CENTER FOR ACADEMIC SUCCESS

The Center for Academic Success offers academic assistance to all students through the use of services including tutoring, workshops, and access to computer based programs. For further information, please visit the Center for Academic Success online or in-person.

INFORMATION LITERACY

Information Literacy is a valuable set of skills that empowers students to become agile information seekers who adapt to changing modes of information delivery and are selective, critical, ethical users of information in all formats. These skills are embedded within course work throughout academic programs. "

Student Participation

Students are expected to actively participate in all their courses throughout each term. Those who fail to do so may be administratively withdrawn from individual courses or the College. Faculty members will include their course-specific policies and procedures in each course syllabus.

Students impacted by illnesses, accidents, or other circumstances that will significantly limit their participation in their courses must notify their faculty members as soon as possible.

Students who are withdrawn for a failure to participate will receive a grade of W, WP or WF for the course(s), depending on whether the student was passing or failing at the time of withdrawal. This may affect the student's financial aid eligibility."