San José State University
Computer Science Department
CS 286:Advanced Topics In Computer Science, Sec-02
Python Data Analytics & Prediction, Fall 2018

Course Information

Instructor: Leonard P. Wesley
Department: Computer Science
College of Science, San Jose State University.
Fall Semester, 2018

Course and Contact Information

Instructor:	Leonard Wesley
Office Location:	MH 212
Telephone:	408.924.5287
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Thursdays 2:00PM – 4:00PM On 9/20, 10/25, and 11/29 Office hours will be from 2:30PM – 4:30PM
Class Days/Time:	Tuesdays and Thursdays 7:30PM – 8:45PM
Classroom:	DH 450
Prerequisites:	Completion of an introductory course in Python 2.7 or 3.x such as CS 286 (Python for non CS majors) or equivalent, and completion of CS 146 or equivalent, or CS 151 or equivalent, or instructor consent.

Course Description

Computing topics of current interest in industrial practice. Emphasis on effective use and integration of software/hardware. Different topics may be offered at different times in a short-course lecture/lab format and may be repeated for credit.

Expanded Course Description

The Fall 2018 CS286-02 Advanced Practical Computing Topics will be a fast-paced introduction to the data analysis process, quantitative and qualitative data analysis

methods & techniques, and selected machine learning and prediction algorithms. Instruction will be provided using Python and Python related packages such as Pandas, numpy, scipy, and matplotlib. Additional Python and statistical packages, such as SAS, Minitab, Stata might be required or suggested as appropriate. Students that know R are perfectly welcome to use it for statistical calculation exercises, however instruction in R and how to use it will not be provided. Students should have access to sufficient computational resources to complete assigned programming exercises and projects.

Students enrolled in this course will be expected to have the equivalent knowledge and programming skills gained after successfully completing (i.e., B or better) an introductory Python programming class, data structures class, and/or object oriented programming. Students completing this course will have a sound understanding of how Python can be used to carry out data analysis and to use selected machine learning algorithms to help make predictions in life science and business intelligence contexts.

Learning Outcomes

Upon successful completion of this course, students will:

1. SLO-1 Be familiar with the theory, practice, and application of data analysis and selected machine-learning (ML) prediction algorithms.
2. SLO-2 Have significant knowledge of data analysis topics and tasks such as
    a. Numerical data type identification, visualization
    b. Data cleaning, correction, and imputation
    c. Numerical Summary Measures, Probability Distributions, Sampling Distributions, Confidence Interval
    d. Hypothesis formation and testing, ANOVA, and nonparametric methods
3. SLO-3 Know how to use selected supervised machine-learning (ML) algorithms to make predictions. Examples such as
    a. Regression-Based (Linear, Logistic, Ridge, PLS)
    b. Classification-Based (SVM, Random Forest)
    c. Statistical-Based (Naïve Bayes, Decision Trees)
    d. Generalization (L1 & L2)
    e. Assessing and comparing the power of linear-based ML algorithms
4. SLO-4 How to apply the theory and best practices of data analytics and ML to real-world applications, and interpret results.

Each SLO above corresponds to a learning module that is described in the course calendar below. That is, there are four (4) learning modules that cover the SLOs described above.

Required Texts

All required text, publications, reference material, and so forth will be provided to the class.

Other Optional Reading Material

Python Data Analytics by Fabio Nelli, Publisher APRES Media LLC, 2015, ISBN-13 (pbk): 978-1-4842-0959-2, ISBN-13 (electronic): 978-1-4842-0958-5

Python for Data Analysis by Wes McKinney, 2013, Publisher O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, ISBN: 978-1-449-31979-3

Additional optional reading material will be provided as appropriate.

Course Requirements and Assignments

Course Logistics

Students should expect to spend approximately nine (9) hours per week (on average) outside of the classroom preparing for and completing the assigned course work. This includes reading papers, viewing videos as appropriate, completing homework and programming exercises, and so forth. The amount of time that a studentactually spendswill depend on individual skills and the timeallocated to the course. The nine (9) hours per week estimate is based on previous experiences of the instructor and students. So please plan and schedule accordingly.

Previously, students have asked for special exception to policies and procedures for this course. An example includes asking the instructor for extra assignments or work to help improve a grade. Even if such a request is reasonable in the opinion of the instructor, no exception will be given to a student unless it can be made available to the entire class, AND does not constitute significant extra work on the part of students, instructors, graders and so forth. Students should have no concern that other students will receive special exceptions that will not be available to the entire class.

NOTE: University policy F69-24 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." However, attendance will be required in order to complete and submit many in-class exercises, quizzes, and exams.

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.

Quizzes and Exams

There will be three quizzes, one midterm and a final exam all of which will count toward the final grad as specified in the “Grades” section below. During quizzes and exams, communication with other individuals via any means is strictly prohibited without the express permission of the instructor. Violations will be met with the full impact of SJSU’s academic integrity policy and procedures.

Projects

Several life science, business intelligence, and finance related data analytics and prediction project topics will be described near the start of the course. Projects will involve applying the skills and knowledge learned in the course to the project. Teams of 2-3 students will be formed to work on a selected project topic. Teams will be required to submit a project proposal before starting on a project, and submit a project report along with working code at the end of the course. Individual student scores on a project will be determined by the content and quality of the contribution of each student toward the project. The score on the course project and project presentation will count toward the final grad (percentage wise) as specified in the “Grades” section below.

Reading, Homework, Programming, In-Class Exercises, Participation Assignments

Graded reading, homework, programming, and class participation and brief course feedback assignments will be given almost weekly, and will count toward the final grade. There will be 4 In-class Exercise sessions. These will typically involve forming teams of 2-3 students that work on assigned programming exercises in the classroom. They provide an opportunity to get started on homework programming assignments that are to be submitted on a designated due date. Participation is mandatory, and scores will count toward final grade.

Computational Resources

Students are required to make sure that they have access to sufficient UNIX, Windows, or Mac based computational resources (e.g., computers and software) to carryout assignments in the course. An attempt to offer the course in a classroom with sufficient computation resources will be made by the department to support classroom instruction and demonstrations. However, students should be prepared to bring their portable laptops to class.

Tentative course calendar of assignment due dates & exam dates:

(Please note that course calendar below, and its content is "subject to change with fair notice")

Week and Class Mtg #	**Tue**	**Thur**	**Module # & Name**	**TOPIC**	**Assignment** **See Canvas For Module & Weekly Assignment Details and Due Dates**
Week 1 Class Mtgs 1 & 2	8/21	8/23	#1 Intro To Data Analytics	8/21: Intro To Course: -Topics, learning objectives, course logistics, Instructor background - Greensheet 8/23: - Intro/Overview of Data Analytics - Definitions: descriptive, inferential, and predictive analytics - Data Analytics process/steps, best practices, applications	Learning Module #1 Week #1 Project Team Formation By Instructor
Week 2 Class Mtgs 3 & 4	8/28	8/30	#1 Intro To Data Analytics #2 Data Analytics Tasks	8/28: -Introduction To Data Analysis (cont.) 8/30: - Numerical Data Types - Visualization	Learning Module #1 Week #2 (8/28) Learning Module #2 Week #2 (8/30) August 31st Last Day To Drop Classes Project Teams Considering Possible Projects
Week 3 Class Mtgs 5 & 6	9/4	9/6	#2 Data Analytics Tasks	9/4: - Data cleaning, correction, imputation, normalization 9/6: - **In-Class Exercise 1** Topics Covered 8/30 – 9/4	Learning Module #2 Week 3 Project Proposals Due 9/4

4	9/11	9/13	#2 Data Analytics Tasks	9/11: - Numerical Summary Measures, Probability Distributions, Sampling Distributions 9/13: - Confidence Intervals - Hypothesis formation and testing	Learning Module #2 Week 4
5	9/18	9/20	#2 Data Analytics Tasks	9/18: - **Quiz 1 (~35 mins): Covers Topics Week 1 thru Week 4** - ANOVA 9/20: - ANOVA (cont.) - Nonparametric methods	Learning Module #2 Week 5
6	9/25	9/27	#2 Data Analytics Tasks #3 ML and Prediction	9/25: - Nonparametric methods (cont.) - Deciding sample size 9/27: - Linear Regression	Learning Module #2 Week 6 (9/25) Learning Module #3 Week 6 (9/27)
7	10/2	10/4	#3 ML and Prediction	10/2: - **In-Class Exercise 2** Topics Covered 9/11 – 9/27 10/4: - **Midterm (Full period): Covers Topics Week 1 thru Week 6**	Learning Module #3 Week 7
8	10/9	10/11	#3 ML and Prediction	10/9: - Logistic Regression - Ridge Regression 10/11: - SVM (SVC, SVR)	Learning Module #3 Week 8

9	10/16	10/18	#3 ML and Prediction	10/16: - **Quiz 2 (~35 mins): Covers Topics Week 7 thru Week 8** - Random Forest 10/18: - Random Forest (cont.) - Naïve Bayes - Decision Trees	Module #3 Week 9
10	10/23	10/25	#3 ML and Prediction	10/23: - Generalization (L1 & L2) - Power of Linear ML, VC Dimension 10/25: - Power of Linear ML, VC Dimension (cont.)	Module #3 Week 10
11	10/30	11/1	#3 ML and Prediction #4 Applied Data Analytics	10/30: - **In-Class Exercise 3** Topics Covered 10/2 – 10/25 11/1 - Real World Data Analysis Examples o Predicting inhibitors of mutant BRAF proteins that cause melanoma cancer	Module #3 Week 11 (10/30) Module #4 Week 11 (11/1)

12	11/6*	11/8	#4 Applied Data Analytics	11/6:
				- Real World Data Analysis Examples
				  o Predicting onset of Alzheimer's Disease
				(NOTE: Instructor will be on travel. This lecture will be a reading assignment. No class meeting)*
				
				11/8:
				- **Quiz 3 (~35 mins): Covers Topics Week 9 thru Week 11**
				- Real World Data Analysis Case Study
				  o Predicting onset of Alzheimer's Disease (Cont.)
					Module #4 Week 12

13	11/13	11/15	#4 Applied Data Analytics	11/13: - **In-Class Exercise 4** (Work on Team Projects, Q&A) 11/15: - Real World Data Analysis Case Study o Predicting customer next purchase	Module #4 Week 13
14	11/20	11/22	#4 Applied Data Analytics	11/20: - Real World Data Analysis Case Study o Predicting object identification 11/22: - **Thanksgiving Holiday**	Module #4 Week 14

15	11/27	11/29	#4 Applied Data Analytics	12/6: - Real World Data Analysis Case Study o Predicting multifocal disorders 12/8: - Review For Final Exam - Jeopardy Data Analytics Game w/ Prizes	Module #4 Week 15
			Final Project Code and Project Report Due To Canvas December 13, 2018 By 11:59PM **Final Exam** December 13, 2018, DH 450, 7:45PM to 10:00PM		

SCHEDULE FOOTNOTES:
NONE AS OF AUGUST 2018

Grades *

WRITTEN HOMEWORK (10 at 10 points each)	100 pts
QUIZZES (3 at 50pts each)	150 pts
MIDTERM	100 pts
IN-CLASS EXERCISES (4 at 50pts each)	150 pts
WEEKLY COURSE FEEDBACK (14 at 5pts each)	70 pts
PROGRAMMING ASSIGNMENTS (5 @ 40pts each)	200 pts
FINAL EXAM	200 pts
FINAL PROJECT REPORT & CODE	200 pts
Total Course Points	= 1,170 pts Total

* The total points for each category might change depending on the number of project teams and assignments. The instructor reserves the right to adjust, with sufficient advanced notice, the above point distributionby ± 5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Grading Percentage Breakdown (NOTE: Ranges might change if point totals change)

Percentage of Total Pts	Pts	Letter Grade
96.66% and above	≥ 1,131	A+
93.33% - 96.65%	≥ 1,092- 1,130	A
90% - 93.32%	≥ 1,053 – 1,091	A-
86.66% - 89.99%	≥ 1,013 – 1,052	B+
83.33% - 86.65%	≥975 – 1,012	B
80% - 83.32%	≥ 936 – 974	B-
76.66% - 79.99%	≥ 897 – 935	C+
73.33% - 76.65%	≥ 858 - 896	C
70% - 73.32%	≥ 819 – 857	C-
66.66% - 69.99%	≥ 778– 818	D+
63.33% - 66.65%	≥ 741 – 777	D
60% - 63.32%	≥ 702 - 740	D-
Below 60%	**< 702**	**F**

HOW TO CALCULATE/ESTIMATE YOUR GRADE

If students would like to calculate their numeric grade percentage, the formula is as follows:

Numeric CS 286 Grade Percentage =

$$\frac{Total\ points\ from\ assignments}{Total\ course\ points} x\ 100\%$$

There is no guarantee that grades will be curved. If so, it will be done at the end of the semester. The instructor is already aware that graduate students need to maintain an overall GPA of B or better. Just because a student NEEDS a particular grade doesn't mean that the instructor will automatically GIVE the student that grade. Students must EARN a passing grade based on submitted and evaluated course work.

Extra credit options, if available

There are no extra credit assignments in this course except for completing designated "Advanced" assignments. However, homework assignments and exams might contain extra credit options.

Penalty for late or missed work

Late assignments will receive a 25% deduction for every 24hr period the submission is late. There will be partial credit for assignments.

Receiving An Incomplete (I) Grade

Receiving a grade of Incomplete (I) is not automatic. Students must complete at least 80% of course assignments by the end of the semester to be eligible to receive a grade of incomplete. Students must also provide documentation to support the reason for the request to receive an Incomplete grade. The instructor has the final decision to give an Incomplete grade. If the instructor agrees to give a student an Incomplete grade, the instructor will enter the remaining work to be completed as part of the PeopleSoft grade submission process.

Grade Change Policy

It is a university policy that course grade changes must be made within one semester from the end of the course. Requests for exceptions to this policy must be accompanied with a documented and compelling reason.

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 at http://www.sjsu.edu/gup/syllabusinfo/. Make sure to review these policies and resources.