

San José State University
Computer Science Department
CS 185C: Advanced Topics In Computer Science, Sec-02
Data Analytics & Prediction, Spring 2020

Course Information

Instructor: Leonard P. Wesley

Department: Computer Science
College of Science, San Jose State University.
Spring Semester, 2020

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, Except on 2/27, 3/9, 3/30 when office hours will be from 12noon to 1:30PM
Class Days/Time:	Tuesdays and Thursdays 4:30PM – 5:45PM
Classroom:	MH 222
Prerequisites:	Completion of an introductory course in Python 2.7 or 3.x such as CS 22B (Python for non CS majors) or equivalent, and completion of CS 146 or equivalent, or CS 151 or equivalent, or instructor consent.

Official SJSU Catalogue 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 Spring 2020 CS 185C-02 Advanced Practical Computing Topics will be a fast-paced introduction to the data analysis process, quantitative and qualitative data analysis

methods, techniques, theory, 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 for students to learn how to use in order to complete homework, in-class exercises, and exam assignments. 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, exams, 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, or object oriented programming. Students completing this course will have a sound understanding of how conduct complex data analysis and use selected ML algorithms to help make predictions in life science, business intelligence, or similarly complex domains of application using Python.

Learning Outcomes

Upon successful completion of this course, students will have:

1. SLO-1 INTRODUCTION: Developed familiarity with the best practices, theory, methods, and application of data analytics. Develop basic skills with applying the best practices of data analysis methods and techniques.
2. SLO-2 DATA ANALYTICS: Obtain significant knowledge about data analysis topics and tasks such as
 - a. Numerical data type identification and data 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 MACHINE LEARNING (ML): Learn about the theory and develop the skills to use selected supervised ML algorithms to make predictions. Examples include
 - a. Regression-Based (Linear, Logistic)
 - 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 APPLICATION: The ability to apply the theory, practice, methods and above algorithms of data analytics and ML to real-world applications in the life science and business intelligence domains, and to be able to interpret and analyze 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 corresponds to each of the SLOs described above.

Required Texts

All required text, publications, reference material, and so forth will be provided to the class. Example textbooks from which material will be used and provided to the class include but is not limited to are

1. Dean Abbott (2014) **Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst**, John Wiley & Sons, Inc, ISBN: 978-1-118-72796-6
2. Marcello Pagano and Kimberlee Gauvreau (2000) **Principles of Biostatistics**, Brooks/Cole, Duxbury is an imprint of Brooks/Cole, a division of Thomson Learning, ISBN 0-534-22902-6
3. Thomas A. Runkler (2012) **Data Analytics Models and Algorithms for Intelligent Data Analysis**, Springer Vieweg, ISBN 978-3-8348-2588-9

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.

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.

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 student actually spends studying and completing course work will depend on individual skills and the time that the student actually allocates 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, some students have asked for special exceptions 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 the same opportunity 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 made available to the entire class.

NOTE: University policy ([F69-24](#)) 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. Should students miss or leave early from one or more classes, students are responsible for knowing and understanding any and all course subject matter, assignments, exercises, instructions and so forth that are presented or discussed during official scheduled class time.

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/Greensheet.

Quizzes and Exams

There will be three quizzes, one midterm and a final exam all of which will count toward a student’s final grade 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, or 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. Projects in this course will be individual (not team) projects. Project scores will count toward the final grade as specified in the “Grades” section below.

In-Class Exercises

There will be four in-class exercises where groups of two to four will be formed to work on an assigned exercise. In-class participation is mandatory, and an attendance sign-up sheet will be passed around to verify participation. The assigned exercises are intended to reinforce learning and understanding of previous lecture, homework, and programming assignment subject matter by providing hands-on experience with completing the provided assignment. A supplement document named “In-Class Exercise Procedure.pdf” is available on Canvas in the same location as the course Syllabus/Greensheet. The “In-Class Exercise Procedure.pdf” document describes the general organization of all in-class exercise assignments as well as the procedure for completing and submitting all in-class exercises. The “In-Class Exercise Procedure.pdf” document should be treated as part of the Greensheet for this course.

Reading, Homework, Programming, Participation Assignments

Graded reading, homework, programming, class participation and brief course feedback assignments will be given almost weekly. All graded assignments will count toward a student’s final course grade.

Questions and Regrade Requests

All questions about grading and re-grade requests must be presented to the instructor within two weeks from the date that graded assignments, exercises, and exams are returned to the class or by the last day of instruction for the semester (whichever is sooner). Assignments, quizzes, and exams will typically be returned (i.e., posted) to Canvas, or manually handed back in class. General questions about the topics covered in assignments, exams, and course are permissible at any time.

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 #	Thur	Tue	Module/SLO #	TOPIC	Assignment See Canvas For Module & Weekly Assignment Details and Due Dates
Week 1 Class Mtgs 1 & 2	1/23	1/28	#1 INTRO	1/23: Intro To Course: -Topics, learning objectives, course logistics, Instructor background - Greensheet 1/28: - Data Analytics Software Apps/Tools - Intro/Overview of Data Analytics - Definitions: descriptive, inferential, and predictive analytics - Data Analytics process/steps, best 1 practices, applications	Module #1 Week #1
Week 2 Class Mtgs 3 & 4	1/30	2/4	#1 INTRO	1/30: - Numerical Data Types - Visualization 2/4: - Numerical Data Types - Visualization - Project Description	Module #1 Week #2 February 4th Last Day To Drop Classes

Week 3 Class Mtgs 5 & 6	2/6	2/11	#2 DATA ANALYTICS	<p>2/6:</p> <ul style="list-style-type: none"> - Data cleaning, correction, imputation, normalization <p>2/11:</p> <ul style="list-style-type: none"> - In-Class Exercise 1 Topics Covered 1/23 – 2/6 	<p>Module #2 Week 3</p> <p>Project Selection Due 2/11</p> <p>February 11th Last Day To Add Classes</p>
4	2/13	2/18	#2 DATA ANALYTICS	<p>2/13:</p> <ul style="list-style-type: none"> - Project Descriptions - Numerical Summary Measures, Probability Distributions, Sampling Distributions <p>2/18:</p> <ul style="list-style-type: none"> - Binomial, Poisson, and Normal Distributions 	<p>Module #2 Week 4</p>
5	2/20	2/25	#2 DATA ANALYTICS	<p>2/20:</p> <ul style="list-style-type: none"> - Quiz 1 (~35 mins): Covers Topics Week 1 thru Week 4 <p>2/25:</p> <ul style="list-style-type: none"> - Sampling Distributions - Central Limit Theorem 	<p>Module #2 Week 5</p> <p>Project Selection Due 2/25</p>
6	2/27	3/3	#2 DATA ANALYTICS	<p>2/27:</p> <ul style="list-style-type: none"> - Confidence Intervals (CIs), Two-sided CIs, One-sided CIs, Hypothesis Formation and Testing <p>3/3:</p> <ul style="list-style-type: none"> - In-Class Exercise 2 Topics Covered 2/6 – 2/27 	<p>Module #2 Week 6</p>

7	3/5	3/10	#2 DATA ANALYTICS	3/5: - Hypothesis Formation and Testing (cont.) 3/10: - Midterm (Full period): Covers Topics from Week 1 thru Week 6	Module #2 Week 7
8	3/12	3/17	#2 DATA ANALYTICS	3/12: - Hypothesis Formation and Testing - Two-Sided Tests of Hypotheses - One-sided Tests of Hypotheses 3/17: - Types of Errors - Power - Sample Size Estimation	Module #2 Week 8
9	3/19	3/24	#2 DATA ANALYTICS	3/19: - One-Way ANOVA - Multiple Comparisons 3/24: - Quiz 2 (~35 mins): Covers Topics Week 7 thru Week 8 - Finish ANOVA	Module #2 Week 9
10	3/26	3/31	#3 MACHINE LEARNING	3/26: - In-Class Exercise 3 Topics Covered 3/3 – 3/24 3/31: SPRING BREAK STARTS 3/30	Module #3 Week 10

11	4/2	4/7	#3 MACHINE LEARNING	<p>4/2: SPRING BREAK ENDS 4/3</p> <p>4/7</p> <ul style="list-style-type: none"> - Regression Theory & Concepts - Assumptions - Simple Linear Regression 	Module #3 Week 11
12	4/9	4/14	#3 MACHINE LEARNING	<p>4/9:</p> <ul style="list-style-type: none"> - Multiple Linear Regression Theory & Concepts <p>4/14:</p> <ul style="list-style-type: none"> - Intro to SVM (SVC) 	Module #3 Week 12
13	4/16	4/21	#3 MACHINE LEARNING	<p>4/16:</p> <ul style="list-style-type: none"> - In-Class Exercise 4 (Work on Individual Projects, Q&A) <p>4/21:</p> <ul style="list-style-type: none"> - Complete SVM - Generalization (L1 & L2) Theory & Concepts 	Module #3 Week 13
14	4/23	4/28	#3 MACHINE LEARNING	<p>4/23:</p> <ul style="list-style-type: none"> - Model evaluation <p>4/28:</p> <ul style="list-style-type: none"> - Quiz 3 (~35 mins): Covers Topics Week 9 thru Week 13 	Module #3 Week 14
15	4/30	5/5	#4 APPLICAT- IONS	<p>4/30:</p> <ul style="list-style-type: none"> - Complete Generalization (L1 & L2) <p>5/5:</p> <ul style="list-style-type: none"> - Introduction to case studies of data analytics 	Module #4 Week 15
16	5/7	No Class	#4 APPLICAT- IONS	<p>5/7:</p> <ul style="list-style-type: none"> - Data analytics case studies - Final exam review 	Module #4 Week 16

	<p>Final Project Code and Project Report Due To Canvas May 20, 2020 By 11:59PM</p> <p>Final Exam Wednesday May 13 from 2:45PM to 5:00PM in MH222</p>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

SCHEDULE FOOTNOTES:

NONE AS OF Spring 2020

Grades *

WRITTEN HOMEWORK (6 at 10 points each)	60 pts
QUIZZES (3 at 50pts each)	150 pts
MIDTERM	100 pts
IN-CLASS EXERCISES (4 at 50pts each)	200 pts
WEEKLY COURSE FEEDBACK (14 at 5pts each)	70 pts
PROGRAMMING ASSIGNMENTS (3 @ 40pts each)	120 pts
FINAL EXAM	200 pts
FINAL PROJECT REPORT & CODE	200 pts

 Total Course Points = 1,100 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 distribution by ± 5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Grading Percentage Breakdown		
Percent of Total Points	Points	Letter Grade
96.66%	\geq 1063	A+
93.33%	\geq 1027	A
90.00%	\geq 990	A-
86.66%	\geq 953	B+
83.33%	\geq 917	B
80.00%	\geq 880	B-
76.66%	\geq 843	C+
73.33%	\geq 807	C
70.00%	\geq 770	C-
66.66%	\geq 733	D+
63.33%	\geq 697	D
60.00%	\geq 660	D-
59.99%	< 660	F

(NOTE: Ranges might change if point totals change)

How To Calculate/Estimate Your Grade

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

Numeric Grade Percentage =

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

There is no guarantee that grades will be curved. If so, it will typically 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

There are no pre-planned extra credit assignments in this course. However, homework assignments and exams might, on occasion, contain extra credit options/questions. At times, the instructor might announce the availability of extra exercises or assignments. There is no guarantee that such extra credit exercises or assignments will be offered to the class. If, in the opinion of the instructor, offering such extra credit options will be significantly advantageous to the learning process, they might be offered.

Late Assignment Submission

Late assignments will receive a 25% point deduction of a graded assignment for each 24hr period the submission is late. For example, if an assignment is worth 10 points, and the grade for the assignment is 8/10, and the assignment is submitted one day late, then the point deduction equals 2.5, and the final grade for the assignment is $\text{MAX}(0, 8 - 2.5) = \text{MAX}(0, 5.5) = 5.5$.

Making Up Missed Assignments

An opportunity to makeup missed exams, homework, in-class exercises, programming assignments, and so forth will be provided if and only if verifiable documentation of a compelling reason (e.g., illness, accident, death in the immediate family) for missing the

assignment is provided within one week from the student's ability to return to class. It is the student's responsibility to (1) contact the instructor if an assignment has or will be missed; (2) obtain verification from the instructor that the student will be allowed to make up the assignment, subject to acceptable and verified documentation; and (3) make arrangements with the instructor to submit all missing assignments by the end of the semester.

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 ([S09-7](#)) that "A change of grade request must be submitted by the department office directly to the Office of the Registrar in a timely fashion. Normally, such requests must be received by the drop deadline of the following Spring or Fall semester ... 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.