IST-718 Big Data Analytics Course Syllabus

Professor: Willard Williamson

Professor Email: wewillia@syr.edu

Office Hours:

- TBD – See Blackboard Announcements
- By appointment

Faculty Assistant: Shubham Sharma

Faculty Assistant Email: shubham.syr14@gmail.com

Faculty Assistant Office Hours:

- TBD – See Blackboard Announcements

- By appointment

Tuesday Class: Hinds Hall Room 117, 5 – 7:45 PM

Thursday Class: Hinds Hall Room 117, 5 – 7:45 PM

Class Discussion Board Sign Up Link:

We will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, please post your questions on Piazza.

Find our class signup link at: piazza.com/home/fall2022/ist718

Contents

Course Description:

A broad introduction to analytical processing tools and techniques for information professionals. Students will develop a portfolio of resources, demonstrations, recipes, and examples of various analytical techniques.

Additional Course Description:

This course will prepare you to participate as a Data Scientist on big data and data analytics projects.

Prerequisite / Co-requisite:

Familiarity with command-line interfaces, quantitative skills including statistics, a basic knowledge of linear algebra, basic probability, basic statistics, basic calculus, strong algebra skills, and strong programming skills in Python or some other language. This is not an introductory course; most students who take this course have already taken IST-687 Introduction to Data Science.

If you have never programmed before, you should probably take an introductory programming course first (preferably in python). If it's been a long time since you took calculus, linear algebra, probability and statistics, it's probably not going to be a problem. You need to understand the high-level concept of a derivative and partial derivative, what a matrix is, how to multiply matrices, and basic probability and statistics. We will use the computer to crunch numbers but we need to understand the high-level math concepts in order to understand the data science topics presented in class. It is assumed that all students are excellent with algebra and this topic will not be reviewed. If you don't have any of the math skills, you might consider taking some prerequisite math courses before you take this class. You should have at least a subset of the recommended math skills.

One test for math skills is to look through some of the assigned reading in Introduction to Statistical Learning in R (ISLR) and Deep Learning (DL) textbooks. If you are totally lost and have never seen any of the math before, this course is probably not for you. ISLR is an especially good test because it is the most used textbook in the class and is very representative of the typical mathematical skill level needed to be successful. The Deep Learning book is used for the math review and only used sparingly beyond that. If you can read and understand the ISLR book, you probably have the mathematical maturity needed to be successful in the class.

In past semesters, the students who struggle the most are the students with the least amount of programming experience. The common denominator for students who struggle the most is a lack of programming experience. Unfortunately many students choose to ignore this warning and don't do very well in the class.

Audience:
- Any student in the School of Information Studies

- o Pursuing an MS in the Applied Data Science program
  - o Pursuing a Certificate of Advanced Study in Data Science
- Any student with an interest in big data analytics.

Credits:

This is a 3 credit course

Learning Objectives:
After taking this course, the students will be able to:
- Translate a business challenge into an analytics challenge
- Use linear and logistic regression, decision trees, and neural networks to make predictions
- Use data science to gain actionable insights
- Use Python and Apache Spark to build big data analytics pipelines
- Apply classic and state of the art machine learning techniques
- Explain how advanced analytics can be leveraged to create a competitive advantage

Texts / Supplies – Required:

- **Python Data Science Handbook (PDSH)** by Jake VanderPlas (Free),
  https://jakevdp.github.io/PythonDataScienceHandbook/
- **An introduction to Statistical Learning with Applications in R (ISLR)** by Gareth James, Daniela, Witten, Trevor Hastie, and Robert Tibshirani (Free)
  http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf
  http://fs2.american.edu/alberto/www/analytics/ISLRLectures.html
- **Deep Learning (DL)** by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (Free)
  http://www.deeplearningbook.org/
- **Apache Spark: The Definitive Guide (ASDG)** by Chambers and Zaharia
  https://www.amazon.com/Spark-Definitive-Guide-Processing-Simple/dp/1491912219/ref=sr_1_1?crid=SY2OHTSAXNQB&keywords=apache+spark+definitive+guide&qid=1564328939&s=gateway&sprefix=apache+spark+definitiv%2Caps%2C150&sr=8-1. This is the only book in the course that is not available for free.
- **Open Intro Statistics (OIS)** by David Diez et al., 4th edition
  (https://www.openintro.org/stat/textbook.php, Texts / Supplies– Additional:

Course Requirements and Grading:

Expectations
All assignments (including project code) must run in the designated class runtime environment. If the assignment has a runtime error in the designated runtime environment, a minimum of 50% is deducted from a specific problem's total points. For example, if a problem is worth 20 points, is made up of 5 sub sections, and any one of the sub sections has a runtime error, your maximum score on that problem will be 10 points. We also reserve the right to deduct more

than half if the problem has especially bad runtime problems. The amount of points deducted is based solely on our judgement but starts at 50% off.

Students are required to read and follow the instructions on all assignments. IST-718 is a 700 - level graduate course at a major university. Students at this level should already know to follow instructions without being told.

If you feel your assignment was not graded correctly, you are free to request an assignment regrade. ALL regrade requests must be made in writing as per the policies below.

- I will not answer regrade questions at the start, during breaks, or after class. I would like the chance to thoughtfully consider all regrade requests and I can't do it during short class breaks.
- Note that regrade requests open your assignment up for the entire assignment to be re-evaluated. For example, if I find an error that was previously missed, I will correct the grading by taking points off.
- The main goal of grading is to apply grading fairly and consistently which means applying grading policies the same for all students.
- All regrade requests must be submitted in writing via email. Do not submit regrade requests to the teaching assistants. Teaching assistants will simply forward your regrade requests to me.
- Your regrade request must include the following:
- Email Subject Line: "Regrade Request"
- Include in the Email Body:
  - The specific assignment and question.
  - Explain in detail why you think the grading is incorrect. Some valid and invalid regrade request examples follow.
    - Valid regrade request example 1: "Greeting Professor. The grading feedback cell for homework 1, question 2 in my notebook says I got 5 points off because I didn't answer question 2c. However, I did answer this question – is it possible the grader didn't see my answer?"
    - Valid regrade request example 2: "Greetings Professor. On homework 1, I did not receive 5 points extra credit but I did the extra credit problem. Did you forget to grade the extra credit?"
    - Invalid regrade request example 1: "Greetings Professor. I got 5 points off for a runtime error on homework 1, question 3. The cell in the question indeed has an error. I feel that I should still get full credit." This request will be denied, please see policies above and do not submit regrade requests of this nature.
    - Invalid regrade request example 2: "Greetings Professor. I got 2 points off because I didn't include a plot title and axis labels for problem 3 on homework 3. Please give me full credit and I will be sure to include plot titles and axis labels in the future." This request will be denied, please do not submit regrade requests of this nature.

Students are required to read and understand all homework instructions.

- ▪ Invalid regrade request example 3: "Greetings Professor, I didn't get full credit for question 2. Can you please explain why." This request will be ignored because there is not enough information to answer the question. The assignment number is missing and there is not enough information about what the student didn't understand. Be very specific about what you are asking for.

- My regrade decisions are final. There is no appeal process beyond the professor. Before submitting a regrade request, read the above policies. If your appeal falls under one of the above policies, you will simply be referred to re-read the above policies and the case will be considered closed.
- If you do not understand a regrade decision, you are welcome to discuss the decision **during office hours**. Please do not ask about regrade decisions before class, during class breaks, or after class. In addition, please do not come to office hours to argue that you should get full credit for any of the above policies. For example, don't come to office hours to argue that you should get full credit even though you clearly didn't follow assignment instructions. Your request will be denied.

Assignment / Points Table

Course assignments will consist of 6 equally weighted homework assignments and one group project. The preliminary grading table below describes the points breakdown in more detail.

Assignment	Description	Points
Academic Integrity Quiz / Piazza Registration / Send Github Username	• Students are required to complete an academic integrity quiz at the beginning of the semester. You are allowed to take the quiz as many times as needed in order to obtain a grade of 100%. • All homework questions are answered via the class piazza discussion board. Students are required to register on the course piazza web site. • Homework is typically distributed via Github Classroom. You need to send the FA your github user name.	25
6 Homework Assignments	100 points per homework assignment	600
Project Proposal	A project proposal due approximately mid semester.	50
Project Report	A comprehensive project report due at the end of the semester.	100
Project Code	Well designed, commented, organized code which runs without error. All project code must run on Colab. All data wrangling, inference, and machine learning methods shall be performed in spark. Visualizations can be performed using any desired python package that will run on data bricks. Project code is due at the same time as the project report.	50

Assignment	Description	Points
In Class Project Presentation	• A 20 minute in class project presentation where all students in the project group participate in the presentation. • Individual students will get a zero for their individual grade related to the project presentation if the student arrives late, leaves early, or doesn't participate in the presentation. • Each student on the project team is expected to present a portion of the presentation. Any student who does not present will receive a 0 grade for the presentation. • Students who miss the project presentation by claiming a University attendance exemption will be required to take an extended version of the final exam.	75
Attendance	See course attendance policy elsewhere in this document for more information on how attendance affects grading.	0
Final Exam	Blackboard online final exam. The final exam will not be math based; but rather, test the students understanding of high-level concepts. Throughout the semester, we will emphasize important high-level concepts. The goal of the final exam is to test the students understanding of the high-level concepts. Note that math may come up on the final exam but it will be related to high level understanding. For example, you will not be asked to perform a calculus derivative. However, you may be asked a question that requires that you understand the high-level concept of a derivative. The goal of the final exam is to be challenging to the best students in the class and give the very best students in the class the chance to separate themselves from the rest of the class.	100
Total Points	**Note: Final Average % = Total Earned Points / Total Points**	**1000**

Final Course Grade Table

Note that the A grade range is a little narrower than typical university grading standards which makes it slightly harder to get an A in the course. The typical A grade range is 93 – 100 but in this class an A grade is 95 – 100. However, the A-, B+, B, and below grade ranges are extended. For example, the B grade range extends down to 80%. Please consider the extended grade ranges below before making a decision to drop the class.
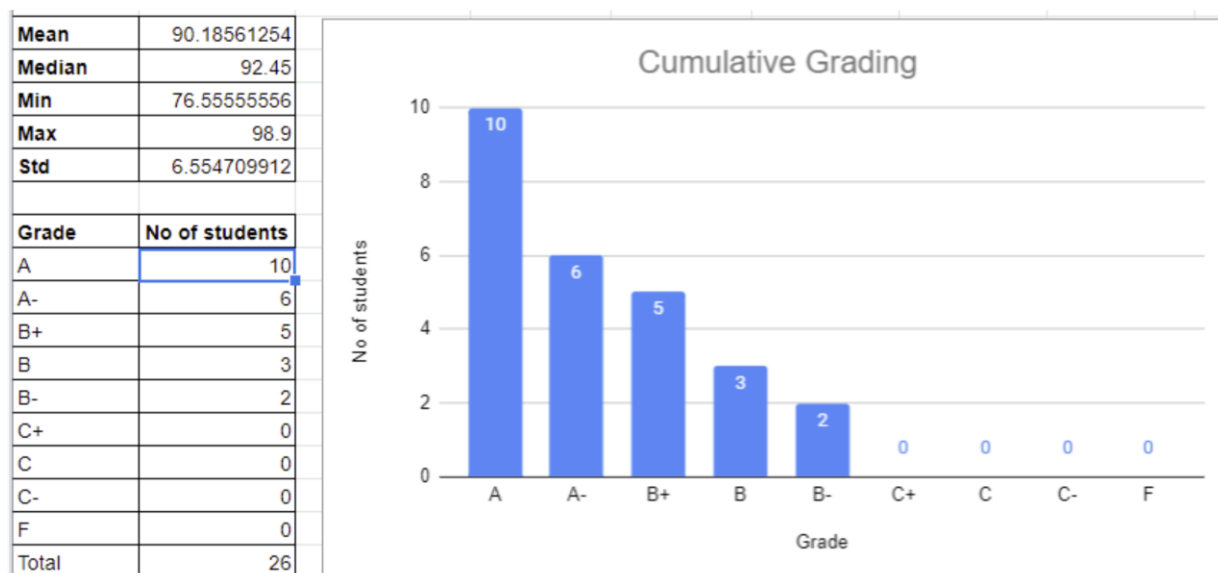
Grades	Grade points /credit	Percentage Range
A	**4.000**	95 – 100
A-	**3.667**	90 – 94
B+	**3.333**	85 - 89
B	**3.000**	80 - 84
B-	**2.667**	75 - 79

Grades	Grade points /credit	Percentage Range
C+	2.333	70 - 74
C	2.000	65 - 69
C-	1.667	60 - 64
F	0	0 - 59

Typical Grading Distribution

I try to distribute grades such that there are a reasonable number of A grades in the class. The number of A grades that is reasonable is determined solely by me at the end of the semester. The table and plot below is a sample grading distribution from IST-718 in the spring semester of 2020. Typically, students on the lowest end of the grading spectrum are found to be in violation of the course academic integrity policy.

IST-718 Spring 2020 Final Grade Distribution

Mean	90.18561254
Median	92.45
Min	76.55555556
Max	98.9
Std	6.554709912

Grade	No of students
A	10
A-	6
B+	5
B	3
B-	2
C+	0
C	0
C-	0
F	0
Total	26

- In the distribution above, 38% of students got A grades.

Course Grade Curve

- I reserve the right to curve final grades. If I do add a curve, it will only improve student grades. I never add a curve which reduces student grades in any way.
- I typically do not curve individual assignment grades.
- I do not artificially decrease grades. If everyone in the class gets an A grade, I will award A grades to everyone in the class. Notwithstanding, if I see an unusually high number of A grades, I may issue more challenging assignments. But if everyone gets A grades, I will not reduce those grades.

- If I do apply a grade curve it will not apply to students involved in academic integrity violations or who have not met the minimum class attendance threshold (see IST-718 Attendance Policy and IST-718 Academic Integrity Policy in this document for more information).
- I may look to see if students attempted extra credit opportunities before applying a grade curve. If students were not interested in improving their grade by doing extra credit throughout the semester, I don't feel that I should give them free extra credit at the end of the semester.
- I may give special grading treatment to students who showed a big improvement throughout the semester and met other requirements outlined above.
- The grade curve is totally at my discretion and can be applied or withheld for any reason I see fit.

Assignment Philosophy

- Homework assignments and the project assignment are designed to get students familiar with typical data science tasks that they will be expected to perform in industry.
- Assignments are challenging from the perspective that students need to read and understand the in class ipython notebooks and then apply what they learned in new / different ways. In other words, we don't provide worked out preparatory examples that are almost identical to the homework assignments. To get an A grade, you must learn the material, think about what you learned, then apply what you learned to solve the homework assignments.
- We find that students often just blindly copy code from the in-class notebooks without understanding how the code works. This is not how you should approach the class. To get an A in the class you will need a thorough understanding of exactly how the in-class notebooks work.
- Assignments stress following instructions and paying attention to detail which are valuable skills in industry.

Homework Grading Philosophy:

- We try to award at least half credit to all assignment questions if the student was kind of on the right track and made a good effort.
- For each problem, we generally take 2 or 3 points off per infraction until we reach 50% off, then stop taking points off as long as the student made a reasonable effort.
- Zero points are awarded for all questions which were not attempted.
- The teaching staff generally doesn't debug your code. If we have runtime problems, we may try and fix it if it's easy to solve. If it takes more than a couple of minutes, we will just start taking off full credit.
- If you have runtime errors, that involve multiple questions, points will be deducted for multiple questions which could result in a very low grade.
- We take points off for not following instructions. Students need to read and understand all assignment instructions.

- Level of effort is considered while grading. We may take points off on a problem if we feel that a student did not similar effort to other students in the class.

Late Days

I realize that unexpected things come up during the semester. To compensate for unexpected life events, all students will get 5 free late days for homework assignments excluding the project. Late days are allotted in whole late days. For example, assuming that homework 1 is due at midnight (12:00 AM) on Tuesday, the student currently has 5 late days remaining, and the assignment is turned in at 12:00 PM on Tuesday (12 hours after the deadline), one late day will be used and the student will have 4 late days left. Late days are always applied in whole day increments.

After all late days are used, 1 letter grade will be deducted for each late day. For example, assume that a student used all 5 late days (0 late days left), turned in an assignment 12 hours late, and got a 95% on the late assignment. The assignment will be graded as 95% – 10% = 85%; one full letter grade would be deducted from the assignment.

Late days cannot be used for the project code, report, or in class presentation.

My advice is to use your late days wisely and don't waste them on trivial matters. Late days are intended to be used for unexpected life events, religious holidays, sports team events, etc.

IST-718 Attendance Policy

- Student attendance does not reduce a student's grade. For example, I don't take point off in any way for missing class.
- Notwithstanding the bullet item above, student attendance is considered when awarding course grade curves. For example, if a grade curve is applied to the course grades, students who have a poor attendance record will not benefit from the grade curve. The determination of a student's attendance record for grading purposes at the end of the semester is determined solely by me. If a student was present for less than N classes (where N is determined solely by me at the end of the semester), then that student will not receive the benefit of a course grade curve. This can sometimes result in missing out on a significant grade boost at the end of the semester. See Course Grade Curve in this document for more information.
- Attendance is only taken during class. If a student does not indicate that they are present during class, then the student is considered absent. I do not allow attendance record updates after class has ended.
- If during the attendance process it is found that a student indicates another student is present when in fact that the other student is not present, the instructor reserves the right to sanction the offending student's grade and / or refer the matter to the University for disciplinary action.
- If students have a valid excuse for missing class like being sick, then I will indicate an excused absence exception in the attendance sheet.

- In cases of extreme absenteeism, I may award the student an F, reduce a student's final grade, or administratively withdraw the student from the course. The definition of extreme absenteeism is determined solely at my discretion.
- **It is always best to email me if you are going to miss class.**
- It is okay to miss the first class of the semester but you must email me if you are going to miss the first class. If you miss the first class and don't email me saying that you intend to be there for the rest of the course, you may be administratively dropped.
- Attendance is mandatory for project presentations. Students who are not present, arrive late, or leave early will get a 0 for their portion of the project presentation. Attendance will be taken at the start and the end of class. I will also verify that all students on the team participate in the presentation.

IST-718 In Class Policies

- It is not okay to have a conversation with other students during lecture. It is rude, disruptive to students who are trying to listen, and disruptive to me.
- It is always okay to ask questions or provide feedback that adds value to the learning process.
- It is always okay to arrive to class late or leave early.
- You are allowed to bring drinks and / or food to class.

How to Succeed in This Course

- Start all assignments early. It's better to have free time at the end of the assignment (so you can work on your project 😊) because you finished it early rather than be trying to work as fast as possible to get the assignment done.
- Starting early allows you to get questions answered which could be blocking progress. You don't want to burn late days waiting to get questions answered.
- Show up regularly for lecture.
- Participate in the piazza discussion forum.
- The course textbooks are excellent, do the assigned reading.
- Don't think of the class as something where you are just completing assignments, try to achieve a high level of understanding.
- Read the assignment instructions thoroughly.

Piazza Discussion Forum

Participation on the piazza discussion forum is highly encouraged. Everyone is allowed to ask and answer questions on piazza. If you know the answer to a question, go ahead and answer the question. However, do not post complete solutions to homework questions on piazza. Do not email the instructor and / or faculty assistants homework questions; but rather, post homework questions on piazza. If you email the instructors directly, you will simply be directed to ask the question on piazza.

IST-718 Academic Integrity Policy

The basic guiding principal of academic integrity in this course is that you must submit work written entirely by you. You are allowed to discuss problems with your class-mates. However, everything you submit must be written by you except for short code snippets you may have obtained from web sites like stackoverflow.com. Code presented in class is always acceptable to copy and submit in homework assignments or in the project. Essentially, you can get ideas for how to solve problems from class-mates, instructors, and the internet. **However, the work you submit in your homework assignments must be written entirely by you.**

- Each academic integrity violation will be evaluated on a case-by-case basis.
- Academic integrity violation grade sanctions can include no action, individual assignment grade reductions, awarding zero grades to individual assignments, or awarding F grades for the entire course. I reserve the right to take any action I see fit.
- I reserve the right to report the student(s) to the university for disciplinary action.
- In cases of an academic integrity violation, I will re-examine all homework assignments for the entire semester for the students in question. I typically find that if I catch students copying code on one assignment, I will also find code copying on previous assignments. Assignment grades can be changed to 0 anytime during the semester, even after an assignment grade is released.
- I reserve the right to withhold students who are involved in academic integrity violations from receiving the benefit of course grade curves in the case that I apply a grade curve.

I am a professional software engineer with over 25 years of experience. This means I have been reading and reviewing code in my professional life for 25 years. After working with a group of software engineers for a short period of time, I can oftentimes tell who wrote a piece of code just by the way it is written. People have unique ways of expressing themselves in code, much like people express themselves with the written word. I am very skilled in identifying unique patterns and signatures in code. If you copy your class-mate's code, I will see that when I grade your homework. Unfortunately, I have identified cases of blatant cheating every semester that I have taught this course. In many cases, I could even tell who provided the code and who received the code based on the patterns and how the software behaved at run time.

As a student in IST-718, you are a senior in a master's degree program at a major research university. To get to this point in your educational career, you have demonstrated a high level of skill and competence in your field. This is not the minor leagues and you are not a student at a community college. Students are expected to know and understand the academic integrity policy.

Allowed Activities
- Discuss approaches on how to solve homework problems.
- Discuss if problems seem like they are correct or not.
- Help a class mate understand a key concept that will help solve a homework or project problem.
- Use web sites like stackoverflow.com to help solve homework and project problems
- Use short code snippets from stackoverflow.com to help solve homework and project problems.

- Help each other with python and spark runtime problems. It's fine to show a class mate your code to get help with a run time problem if you trust that your class-mate will not copy your code and submit it as their own.
- Think of it like this: The goal of the class is for YOU to develop skills to solve data science problems and write code that implements solutions to those problems. You can get ideas from others on how to solve problems but YOU are the one that needs to turn the ideas into code that implements the solution.

Prohibited Activities

- **Submitting code or written answers which were written by someone other than you.**
- Coding solutions to homework problems in pairs or groups then slightly changing the answers across the assignments in the group and submitting. I actually found cases of this every semester I have taught the class. Often the students involved say they are new at programming. This is not an acceptable excuse.
- Copying a complete homework or project solution from the internet.
- Copying homework assignments from a previous semester.
- Submitting project work that you did in another class. If you previously turned in a project for another class, you are not allowed to resubmit it for this class. You are also not allowed to slightly modify a project you submitted for another class and submit it for this class. It is possible that you could perform a major extension to a previous project from another class. Please consult with me to avoid the possibility of academic integrity violations.
- Providing part or all of your completed assignment to another student. In the past, we have had several cases where a student provided a completed assignment to another student in the class. The student who received the assignment then copied the entire assignment and turned it in as their own work. BOTH students received zeros for the assignment. **You are not allowed to provide other students with partially or fully completed homework assignments.**

Note that students are required to complete an ungraded academic integrity quiz on blackboard and obtain 100% on the quiz. Students who do not complete the quiz by the due date will be charged late days until it is complete. Homework assignments will not be accepted and will be counted as late until the academic integrity quiz is complete.

Syracuse University Policies:

Syracuse University has a variety of other policies designed to guarantee that students live and study in a community respectful of their needs and those of fellow students.

University Attendance Policy
The following is the Official University attendance policy. See grading policies above for how attendance ties into grading.

Attendance in classes is expected in all courses at Syracuse University. Students are expected to arrive on campus in time to attend the first meeting of all classes for which they are registered.

Students who do not attend classes starting with the first scheduled meeting may be academically withdrawn as not making progress toward degree by failure to attend. Instructors set course-specific policies for absences from scheduled class meetings in their syllabi.

It is a federal requirement that students who do not attend or cease to attend a class to be reported at the time of determination by the faculty. Faculty should use "ESPR" and "MSPR" in Orange Success to alert the Office of the Registrar and the Office of Financial Aid. A grade of NA is posted to any student for whom the Never Attended flag is raised in Orange SUccess. More information regarding Orange SUccess can be found here, at:
http://orangesuccess.syr.edu/getting-started-2/
Students should also review the University's religious observance policy and make the required arrangements at the beginning of each semester

Diversity and Disability

Information on ensuring that students are aware of their rights and responsibilities in a diverse, inclusive, accessible, bias-free campus community can be found here, at
https://www.syracuse.edu/life/accessibilitydiversity/

Religious Observances Notification and Policy
Steps to follow to request accommodations for the observance of religious holidays can be found here, at: http://supolicies.syr.edu/studs/religious_observance.htm

Orange SUccess
Tools to access a variety of SU resources, including ways to communicate with advisors and faculty members can be found here, at: http://orangesuccess.syr.edu/getting-started-2/

Disability-Related Accommodations:
Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. There may be aspects of the instruction or design of this course that result in barriers to your inclusion and full participation in this course. I invite any student to meet with me to discuss strategies and/or accommodations (academic adjustments) that may be essential to your success and to collaborate with the Office of Disability Services (ODS) in this process.

If you would like to discuss disability-accommodations or register with ODS, please visit their website at http://disabilityservices.syr.edu./ Please call (315) 443-4498 or email disabilityservices@syr.edu for more detailed information.

ODS is responsible for coordinating disability-related academic accommodations and will work with the student to develop an access plan. Since academic accommodations may require early

planning and generally are not provided retroactively, please contact ODS as soon as possible to begin this process.

University Academic Integrity Policy:

Syracuse University's Academic Integrity Policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same work in more than one class without receiving written authorization in advance from both instructors. Under the policy, students found in violation are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered as described in the Violation and Sanction Classification Rubric. SU students are required to read an online summary of the [University's academic integrity](#) expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check-in on [MySlice](#).

Educational Use of Student Work

Student work prepared for University courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing to be enrolled in, courses where such use of student work is announced in the course syllabus.

The professor may use academic work that you complete this semester for educational purposes in this course during this semester. Your registration and continued enrollment constitute your permission.

The professor may also use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identifications.

As a generally accepted practice, honors theses, graduate theses, graduate research projects, dissertations, or other capstone projects submitted in partial fulfillment of degree requirements are placed in the library, University Archives, or department for public reference.

Discrimination or Harassment

The University does not discriminate and prohibits harassment or discrimination related to any protected category including creed, ethnicity, citizenship, sexual orientation, national origin, sex, gender, pregnancy, disability, marital status, age, race, color, veteran status, military

status, religion, sexual orientation, domestic violence status, genetic information, gender identity, gender expression or perceived gender.

Any complaint of discrimination or harassment related to any of these protected bases should be reported to Sheila Johnson-Willis, the University's Chief Equal Opportunity & Title IX Officer. She is responsible for coordinating compliance efforts under various laws including Titles VI, VII, IX and Section 504 of the Rehabilitation Act. She can be contacted at Equal Opportunity, Inclusion, and Resolution Services, 005 Steele Hall, Syracuse University, Syracuse, NY 13244-1120; by email: titleix@syr.edu; or by telephone: 315-443-0211.

Federal and state law, and University policy prohibit discrimination and harassment based on sex or gender (including sexual harassment, sexual assault, domestic/dating violence, stalking, sexual exploitation, and retaliation).

If a student has been harassed or assaulted, they can obtain confidential counseling support, 24-hours a day, 7 days a week, from the Sexual and Relationship Violence Response Team at the Counseling Center (315-443-4715, 200 Walnut Place, Syracuse, New York 13244-5040). Incidents of sexual violence or harassment can be reported non-confidentially to the University's Title IX Officer (Sheila Johnson Willis, 315-443-0211, titleix@syr.edu, 005 Steele Hall). Reports to law enforcement can be made to the University's Department of Public Safety (315-443-2224, 005 Sims Hall), the Syracuse Police Department (511 South State Street, Syracuse, New York, 911 in case of emergency or 315-435-3016 to speak with the Abused Persons Unit), or the State Police (844-845-7269).

I will seek to keep information you share with me private to the greatest extent possible, but as a professor I have mandatory reporting responsibilities to share information regarding sexual misconduct, harassment, and crimes I learn about to help make our campus a safer place for all.

Course evaluations:
There will be an end of course evaluation for you to complete this term. This evaluation will be conducted online and is entirely anonymous. You will receive an official notification in your email account with the evaluation website link and your passcode. Please take the time and fill out this evaluation as your feedback and support of this assessment effort is very much appreciated. The school carefully reviews ratings and comments that you submit, and these factor into decisions about course, program and instructor development.

Use of Blackboard:
This course makes minimal use of Syracuse University's Blackboard system as an online tool. To access Blackboard, [http://blackboard.syr.edu] use your Syracuse University NetID & Password. This specific course will appear in your course list.

To search for answers to your Blackboard questions, visit the Answers self-help knowledge [https://answers.syr.edu/display/blackboard01/Blackboard]. If you have problems logging in or

need assistance with Blackboard, contact the ITS Service Center at: help@syr.edu or 315.443.2677. The Syracuse University Blackboard support team will assist you.

Course Schedule

The following is a preliminary course schedule.

Date	Class Num	Topics	Suggested Reading / Notes	Events
Aug 30 Sep 1	1	**Course overview and python programming:** Anaconda installation, jupyter notebooks, Google Colab, git, Numpy, Pandas, Matplotlib PDSH: Ch 1 PDSH: Ch 2.2 – 2.7	• PDSH Ch 2: Numpy • PDSH Ch 3: Pandas • PDSH Ch 4: Matplotlib Note: There is a lot of assigned reading. Skip sections in the above if you are already familiar. Focus on areas you need work. Familiarize yourself with what's available in the book for future reference.	Homework 0 Released Homework 1 Released
Sep 6 Sep 8	2	**Review:** Probability, statistics, linear algebra.	• DL Ch 2.1 – 2.6: Linear Algebra • OIS Ch 3: Probability • OIS: Section 4.1, 4.3: Random Variables	• Homework 0 Due September 9th for both sections
Sep 13 Sep 15	3	**Introduction to Spark, Hadoop, MapReduce, and YARN**	• ASDG Ch 2: Introduction to Spark • ASDG Ch 3: Spark Tour • ASDG Ch 12: RDD's	• Homework 1 Due • Homework 2 Released
Sep 20 Sep 22	4	**Introduction to Spark DataFrames and Spark ML**	• ASDG Ch 4: Data Frames • ASDG Ch 5: Expressions / Operations • ASDG Ch 6: Data Types	• Project Teams Are Due – Students pick their own teams of 1 to 4. Due September 23rd for both sections.

Date	Class Num	Topics	Suggested Reading / Notes	Events
Sep 27 Sep 29	5	**A Statistical Perspective on Machine Learning:** Introduction to linear regression, mean squared error, reducible and irreducible error, regression vs. classification, parametric vs. non parametric models.	• ISLR Ch 1: Introduction • ISLR Ch 2.1: Statistical Learning • ASDG Ch 24: ML Overview	• Homework 2 Due • Homework 3 Released
Oct 4 Oct 6	6	**Assessing Model Accuracy:** Bias variance tradeoff, confusion matrix, receiver operating characteristic, model selection.	• ISLR Ch 2.2: Model Accuracy • ISLR Ch 3.1: Linear Regression • ISLR Ch 3.2: Multiple Linear Regression • ISLR Ch 5.1: Cross Validation	
Oct 11 Oct 13	7	**Sentiment Analysis Case Study:** Supervised learning, model selection, logistic regression, regularized logistic regression, elastic net regularization, model interpretation	• ISLR Ch 4.1: Classification • ISLR Ch 4.2: Why Not Linear Regression • ISLR Ch 4.3: Logistic Regression • ASDG Pages 459 – 465: Classification / Logistic Regression	• Homework 3 Due • Homework 4 Released
Oct 18 Oct 20	8	**Sentiment Analysis Case Study: Continued**	• ISLR Ch 6: Feature Selection / Regularization	• Project Proposal Due
Oct 25 Oct 27	9	**Course Recommendation Case Study:** Unsupervised Learning, Principal component analysis, K-Means clustering.	• ISLR Ch 10.1: Unsupervised Learning • ISLR Ch 12.2: PCA • ISLR Ch 10.3.1: K-Means	• Homework 4 Due • Homework 5 Released

Date	Class Num	Topics	Suggested Reading / Notes	Events
Nov 1 Nov 3	10	**Course Recommendation Case Study: Continued**	• PDSH Chapter 5: In Depth – Principal Component Analysis • PDSH Ch 5: In Depth – K-Means Clustering	
Nov 8 Nov 10	11	**Predicting Credit Score Case Study:** Decision trees, bagging, boosting, random forests. ISLR chapter 5.2	• ISLR chapter 8: Trees	• Homework 5 Due • Homework 6 Released
Nov 15 Nov 17	12	**Predicting Credit Score Case Study: Continued**	• ASDG p 465 – 469: Trees	
Nov 22 Nov 24	Thanksgiving Break			
Nov 29 Dec 1	13	**Deep Learning**	• DL Ch 6: Feed Forward Networks • ASDG Ch 31: Deep Learning	Homework 6 Due
Dec 6 Dec 8	14	**Project Presentations**	N/A	
Dec 9		Project code and report is due for all class sections		
Dec 12 – Dec 16		**Finals Week** Online final exam may be taken anytime during the final exam period		
Dec 27		**Final grades due from faculty**		