# CIS 8392: Advanced Topics in Big Data Analytics

# Course Information

**Semester**: Fall 2021
**Class Hours**: Wednesday 5:30pm–9:45pm
**Class Location**: TBD
**Instructor**: Dr. Yu-Kai Lin

- Email Address: yklin@gsu.edu (mailto:yklin@gsu.edu)
- Office Location: Room 422, Robinson College of Business (35 Broad St)
- Office Hours: By appointment; via WebEx. Etiquette rules:
    1. Please give me at least a **24 hours' notice**. That is, you cannot schedule an appointment that takes place within 24 hours of sending your email.
    2. In the email, you need to provide at least **3 time slots** that work for you, and they need to be on at least **2 different weekdays** and within **normal business hours**.
    3. I will respond to your email within 24 hours to confirm the appointment or propose a different schedule.

# Course Description

This course will cover various modern topics in big data analytics, including natural language processing (NLP), high performance and interpretable machine learning (ML), deep learning for text/image data, and cloud-based technologies. All of these will be demonstrated using R. In other words, this course will teach students how to do advanced data science in R.

We emphasize active and experiential learning. To that end, the course materials are designed to deliver a lab-oriented learning experience with plenty of demos and exercises. Teaching in this course is often stylized as (and based on) a Kaggle competition so that students learn the strategies and skills in solving real-world end-to-end data science projects.

# Course Objectives

Upon completion of the course, students will be able to:

1. Develop R programs for data science
2. Manage the workflow in big data analytics
3. Communicate data analytics procedures and the findings professionally
4. Articulate techniques in big data analytics
5. Construct high performance machine learning pipelines

6. Apply deep learning methods for unstructured data analytics
7. Discuss the motivations for and techniques in cloud-based big data analytics

# Recommended Textbooks / Supplementary Materials

1. **R for Data Science** (2017; O'Reilly), by Hadley Wickham and Garrett Grolemund. Referred to as **RDS** in Syllabus. Freely available at https://r4ds.had.co.nz/ (https://r4ds.had.co.nz/)
2. **Hands-On Machine Learning with R** (2020; CRC Press), by Bradley Boehmke and Brandon Greenwell. Referred to as **MLR** in Syllabus. Freely available at https://bradleyboehmke.github.io/HOML/ (https://bradleyboehmke.github.io/HOML/)
3. **Explanatory Model Analysis** (2021; CRC Press), by Przemyslaw Biecek and Tomasz Burzykowski. Referred to as **EMA** in Syllabus. Freely available at https://ema.drwhy.ai/ (https://ema.drwhy.ai/)
4. **Deep Learning with R** (2018; Manning Publications), by Francois Chollet and J.J. Allaire. Referred to as **DLR** in Syllabus.
5. **Google Cloud Platform in Action** (2018; Manning Publications), by JJ Geewax. Referred to as **GCP** in Syllabus

# Technology and Software Requirements

Because of the lab-oriented teaching style of this course, **students are required to bring their own laptop to every class**. The computer/laptop should be able to install and execute the following software (all free!):

1. R (https://cran.r-project.org/): R is a popular programming language for data analytics. We will learn basic R programming in our lab meetings. The final project and all lab exercises will be based on R.
2. RStudio (https://www.rstudio.com/products/RStudio/): RStudio is an environment for R programming with many user-friendly features.

# Student Evaluation

Gradable items are listed below. There will be occasional opportunities for extra credits.

- Assignments: 60% (=20%*4; dropping the lowest)
- Course Project: 30%, including:
    - Proposal document and presentation: 5%
    - Comments to other teams' proposal presentations: 5%
    - Responses to peer comments: 5%
    - Project final presentation: 5%
    - Project final submission (report, slides, code, and data): 10%
- Participation: 10%

## Expectations When Evaluating Assignments

Assignments are the most important component in student evaluation. While specific requirements in each assignment will vary, below is a list of common sense rules:

1. You must submit your work to iCollege **before** the deadline.
    - You should plan to submit your work at least one hour before the deadline so as to prevent unanticipated network/laptop issues.
    - Submissions made through email will not be graded.
    - You will lose half of the points for a late submission.
    - If a deadline is 5:30pm, it means 5:30:00.000pm. In other words, if your submission is time-stamped at 5:30pm, it is already past the deadline and will be considered as a late submission.
2. Your work must be original. Do not copy codes from other students or from the internet.
    - Academic dishonesty will not be tolerated.
    - You will lose all points if you share your assignment with others.
    - You will lose all points if your assignment is copied from other people.
3. Read and follow the assignment/project instructions carefully.

## Expectations When Evaluating the Final Project Submission

1. You must submit your work to iCollege by the deadline.
    - You should plan to submit your work at least one hour before the deadline so as to prevent unanticipated network/laptop issues.
    - You will receive 0 point if you submit your project via email.
    - You will lose 5 points for late submission.
2. Your work must be original. Do not copy work from other students or from the internet.
    - You will lose all points if your project is copied from other people.
3. Read and follow the project instructions carefully.

## Expectations When Evaluating Participation

Class attendance, while not mandatory, is expected. The instructor encourages everyone to participate in class activities, discussions, and respond to questions from other students. In evaluating your class participation in discussions, both the quantity and quality of participation is taken into account. Principles for class participation include:

1. Show a respectful and positive attitude towards himself/herself, classmates, and teacher
2. Assist peers during lab sessions
3. Work with others in paired/group-based exercises
4. Contribute to classroom discussion
5. Attend classes and **focus on class work** (that is, not using social media, sending emails/texts, or doing anything irrelevant to class activities)

I will take notes whenever you violate any of the above principles, **especially #5**.

# Final Grade

The final letter grade is based on **total points** you earned in the course. The following grading scale will be used to translate total points to final grade:

| Total Points | Grade | Total Points | Grade |
|---|---|---|---|
| >= 97 | A+ | 77-79.9 | C+ |
| 93-96.9 | A | 73-76.9 | C |
| 90-92.9 | A- | 70-72.9 | C- |
| 87-89.9 | B+ | 60-69.9 | D |
| 83-86.9 | B | < 60 | F |
| 80-82.9 | B- | | |

Please note that **iCollege does not support *take-3-out-of-4* calculation for the lab assignment grades**. As such, the grade percentage numbers on iCollege are not useful when it comes to our final grading.

# Course Outline

| Wk | Date | Topic | Reading | Assignment | Project |
|---|---|---|---|---|---|
| 1 | Oct 13 | R for Data Science - part 1 | RDS Ch1-16 | A1 assigned | |
| 2 | Oct 20 | R for Data Science - part 2 | RDS Ch17-30 | A1 due | Setup team/topic |
| 3 | Oct 27 | Explainable AI | EMA Ch5-20 | A2 assigned | Proposal due |
| 4 | Nov 3 | High Performance ML | MLR Ch1-3,11,12 | A2 due | Peer comments due |
| 5 | Nov 10 | Deep Learning for Images | DLR Ch5-6 | A3 assigned | Responses due |
| 6 | Nov 17 | Deep Learning for NLP | DLR Ch1-4 | A3 due | |
| 7 | Dec 1 | Cloud-based AI Systems | GCP Ch14-21 | A4 assigned | |
| 8 | Dec 8 | Project Presentation | | A4 due | Final submission |

The course outline provides a general plan for the class. However, the plan is subject to change to accommodate students' learning progress and unexpected events. All changes to the outline will be updated and posted in iCollege.

# Some R Resources

In addition to the class handouts, there are a number of good resources available on the web for both learning R and seeking answers to questions about how to accomplish various tasks.

- The official intro, "An Introduction to R", available online in HTML (https://cran.r-project.org/doc/manuals/r-release/R-intro.html) and PDF (https://cran.r-project.org/doc/manuals/R-intro.pdf)
- RStudio Cheat Sheets (https://www.rstudio.com/resources/cheatsheets/)

- R Reference Card (https://cran.r-project.org/doc/contrib/Short-refcard.pdf)
- Thomas Lumley, “R Fundamentals and Programming Techniques” (a large PDF slide deck (http://faculty.washington.edu/tlumley/Rcourse/R-fundamentals.pdf))
- Online resources such as Stack Overflow (http://stackoverflow.com/questions/tagged/r) and the R-help Mailing List (https://stat.ethz.ch/mailman/listinfo/r-help)
- RStudio Documentation (https://support.rstudio.com/hc/en-us/categories/200035113-Documentation)
- R and Data Mining: Examples and Case Studies (http://www.rdatamining.com/docs/r-and-data-mining-examples-and-case-studies)
- … and so much more (http://stats.stackexchange.com/questions/138/resources-for-learning-r)

# Course and University Policy Statements

## Syllabus Change Policy

This syllabus provides a general guideline for the conduct of this course. However, deviations will be necessary. Updates will be given during the semester and posted online through iCollege.

## Class Website on iCollege

This website includes a copy of this syllabus (and any subsequent updates or changes), lecture slides, lab notes, readings, and information about the final project. It is your responsibility to check this website frequently for announcements and updates. Copies of class handouts and presentation slides will be posted on the class website before each week’s classes (if not earlier). You may find it helpful to use these to take notes during class.

## Unauthorized Public Posting and Distribution of Course Materials

GSU policy prohibits students from posting instructor-generated materials on external sites.

The selling, sharing, publishing, presenting, or distributing of instructor-prepared course lecture notes, videos, audio recordings, or any other instructor-produced materials from any course for any commercial purpose is strictly prohibited unless explicit written permission is granted in advance by the course instructor. This includes posting any materials on websites such as Chegg, Course Hero, OneClass, Stuvia, StuDocu and other similar sites. Unauthorized sale or commercial distribution of such material is a violation of the instructor’s intellectual property and the privacy rights of students attending the class, and is prohibited.

## Class Attendance

Excused absences include university-sponsored events, legal obligations, and religious observances (see http://codeofconduct.gsu.edu/ (http://codeofconduct.gsu.edu/) ). These absences will be accommodated in a way that does not arbitrarily penalize students who have a valid excuse. Consideration will also be given to students whose dependent children experience serious illness.

When seeking excused absences for medical reasons, students should submit documentation to https://deanofstudents.gsu.edu/student-assistance/professor-absence-notification/ (https://deanofstudents.gsu.edu/student-assistance/professor-absence-notification/) . Professors will then be notified by the Dean of Students of any excused absence without the need to manage medical information individually.

Regardless of whether an absence is excused or unexcused, the student is responsible for making up all work that is missed. No make-up assignments will be given except for exceptional cases that are approved by the instructors or university policy.

# Disabilities or Special Needs

Students who wish to request accommodation for a disability may do so by registering with the Office of Disability Services. Students may only be accommodated upon issuance by the Office of Disability Services of a signed Accommodation Plan and are responsible for providing a copy of that plan to instructors of all classes in which accommodations are sought. Please inform the instructor if you have a disability or special need that requires accommodation.

# GSU Academic Honesty Policy

Students may have general discussions about assignments with fellow classmates, but each student must develop his or her solution to each weekly project/assignment. It is each student's responsibility to keep his/her own work secure. DO NOT share computer files of project/assignment with classmates. Failing to adequately protect one's work does not relieve the student from academic dishonesty charges.

University regulations will be enforced regarding dishonorable or unethical conduct (Cheating, Plagiarism, Falsification, Unauthorized Collaboration or Multiple Submissions). The penalties for incidents of academic dishonesty can lead to expulsion from the University (see http://codeofconduct.gsu.edu/ (http://codeofconduct.gsu.edu/) ). In this class, there will be ZERO TOLERANCE for dishonorable or unethical conduct. Electronic or physical sharing of answers will be considered cheating and will not be tolerated.

If a student is charged with academic dishonesty, for each charge, a zero (0) will be given for the assignment, a minimum of point equivalent of one final grade (i.e. B- to a C-) will be deducted from the final course total points and a written Notice of Academic Dishonesty will be given to the Dean's office. The student will also receive a copy of the notice.

# Grade Appeal and Reassessment

It is important to recognize that a grade reflects others' judgment of your work. In this sense, some grading is inevitably subjective. Of course, any grade you receive on an assignment is subject to appeal. However, score changes are at the discretion of the instructor. It is important to understand that your score may go up or down based upon a complete review of the work in question. It is usually the case that changing a few points on an assignment rarely makes a difference in the final grade. Time is much better spent discussing and clarifying the information content presented in the course. However, if you would like to appeal your grade, you must submit the appeal in writing to me within one (1) week of receiving the graded work.

If a request is made for any assignments to be reassessed, please recognize that it will be possible to retain, gain, or lose points in the reassessment process. Make any reassessment requests by email within one (1) week of grading. Please make a follow-up appointment (preferably email) to meet the instructor during office hours for review of the results of any reassessment.

# Class Disruption

Do not disrupt, distract or prevent others from learning by arriving late, leaving early or failing to turn off all electronic devices during the scheduled class. (Laptop computers used for taking notes and conducting in-class exercises being the exceptions to this rule.)

# Student Feedback to Instructor and Continuous Improvement

I am committed to continuous improvement of my teaching. Please feel free to speak with me about any aspect of the course (both positive and negative). I will consider your feedback and suggest carefully to hopefully make this course a positive and effective learning experience for you.

Your constructive assessment of this course plays an indispensable role in shaping education at Georgia State. Upon completing the course, please take the time to fill out the online course evaluation.