Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2021

Class Information

Course Title:	Applied Data Mining (index#05521)
	Economics 422/Section 01
Instructor:	Dr. I-Ming Chiu
Office:	ARMITAGE 328
	Phone (856) 225 6012
E-mail address:	ichiu@camden.rutgers.edu
Class Meeting:	Remote Teaching via Zoom (11:15 am-12:30 pm, Tuesday/Thursday)
Office Hours:	12:45-1:45 pm, Thursday or by appointment

Course Description: We are living in an era where new data are being fast produced from all kinds of human activities that include our daily emails, social network postings, and transactions on S&P 500 etc. How do we utilize these high-volume data? The answer is "Data Mining". Data mining is a computing process of using mathematical and statistical algorithms to uncover patterns from huge data and transform them into decision-making information and new knowledge. A variety of data mining algorithms, supervised and unsupervised, will be introduced and the topics include linear regression model, logistic regression model, KNN, Support Vector Machines, Naïve Bayes, Association Rules, and K-Means Clustering, etc. Each topic covered is accompanied with a case study, so students will gain many hands-on learning experiences. The ultimate goal in this course is to equip students with modern data analytical tools, which has a high demand in the job market. Please click the link below to find out the potential reward for data science related jobs:
https://www.glassdoor.com/Salaries/us-data-scientist-salary-SRCH_IL.0,2_IN1_KO3,17.htm

Main Readings: Brett Lantz, Machine Learning with R, 3rd edition, Packt Publishing, 2019 (ISBN: 978-1788295864). (Book abbreviation: MLR; available at bookstore)

Max Bramer, Principles of Data Mining, 3rd edition, Springer, 2018. (Book abbreviation: PDM)

Frank Emmert-Streib, Salissou Moutari, and Matthias Dehmer, Mathematical Foundations of Data Science Using R, De Gruyter Oldenbourg, 2020 (ISBN: 978-3110564679). (Book abbreviation: MFDS)

The course material will be drawn from various sources. The MLR book is the main source of the practical data mining tools covered in the class. The PDM book explains the algorithms behind data mining

methods. The MFDS book provides the mathematical details of machine learning algorithms.

Other References: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning/with Applications in R, Springer, 2013 [available via school library].

Hadley Wickham & Garrett Grolemund, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly Media, 2017.

Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar, Introduction to Data Mining 2nd edition, Pearson, 2018.

Computing: All the computations will be done using an open source statistical software R. It can be downloaded at http://www.r-project.org. You're encouraged to download and use RStudio at the following site, which is a very user-friendly IDE (integrated development environment) for R. https://www.rstudio.com/products/rstudio/

R Installation: https://www.youtube.com/watch?v=Icawuhf0Yqo (for Mac)
https://www.youtube.com/watch?v=hxj0UG4boGU (for PC)

Class Material: Data, handouts, assignments, and additional readings will be posted on Sakai website: https://sakai.rutgers.edu/portal.

Online Learning: https://www.datacamp.com/ (Learn Data Science online)

Useful Websites: http://www.statmethods.net/ (Computing using R web site)

https://www.analyticsvidhya.com/ (Machine Learning)

Spring '21 Calendar: https://registrar.camden.rutgers.edu/academic-calendar-2020-2021

Grading: **Contributions to Final Grade**

- Attendance	5%
- Take-home problems & DataCamp courses	40%
- Midterm Exam	25%
- Final Exam or Project	30%
- Participation (extra credit)	5%

Grading Policy: Term grades will be based on the final distribution of the above grading weights.

Exam Preparation: The exam questions will be drawn from three sources: (i) homework assignments, (ii) course lectures, and (iii) reading material.

Class Participation: Class attendance is essential for learning achievement. When missing a class, it would cost you more time to learn on your own. I strongly recommend the following steps for your successful learning: (1) attend every class and take notes; (2) review everything you learn from the class immediately, never put it off; (3) ask questions and participate in class discussions.

Academic conduct: Make up exams will be given **only upon prior notice**. I request prior knowledge of any expected absence from an exam. If this is not feasible, you can document a valid reason for missing the exam. Unexcused absence on any exam will result in a grade of zero. Dishonesty in seeking an excused absence or in the examination process will result in a grade of zero on the exam involved and in university discipline. More detailed information can be found at the following site: https://academicintegrity.rutgers.edu.

Disability Services: Rutgers University welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student with a disability must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation:

https://ods.rutgers.edu/students/documentation-guidelines. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with a Letter of Accommodations. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. To begin this process, please complete the Registration form at https://webapps.rutgers.edu/student-ods/forms/registration.

Here is the link to the Office of Disability Service:
https://success.camden.rutgers.edu/disability-services

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2021

Course Outline:

Topic 1	Mathematical and Statistical Fundamentals
Topic 2	Introduction to Computing Using R
Topic 3	What is Data Mining? The Familiar Linear Regression Model
Topic 4	K Nearest Neighbors (KNN)
Topic 5	Naïve Bayes
Topic 6	Decision Trees
Midterm Exam	Date: TBA in the class
Topic 7	Logistic Regression Model
Topic 8	Support Vector Machines
Topic 9	Association Rules (aka Market Basket Analysis)
Topic 10	Text Mining
Topic 11	Linear Model Selection and Regularization
Topic 12	Introduction to Neural Networks
Final Exam/Project	Details will be explained in the class meeting