

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2023

Class Information

- Course Title:** Applied Data Mining (index#: 16049 & 16050)
Economics 50:220:422/Sec 92 & Psychology 56:830:678/Sec 92
- Instructor:** Dr. I-Ming Chiu
- Office:** ARMITAGE 435
Phone (856) 225 6012
- E-mail address:** ichiu@camden.rutgers.edu
- Class Meeting:** virtual meeting via Zoom, 6:00-8:50 pm, Tuesday
- Office Hours:** 3:30-4:30 pm, Tuesday (in-person) 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:** Fred Nwanganga & Mike Chapple, [Practical Machine Learning in R](#), 1st edition, Wiley Publishing 2020 (book abbreviation: PMLR). Amazon link: [Click here](#). Library link: [Click here](#).
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, [An Introduction to Statistical Learning/with Applications in R](#), 2nd edition, Springer 2021 (book abbreviation: ISLR). Amazon link: [Click here](#). Library link: [Click here](#).
- Max Bramer, [Principles of Data Mining](#), 4th edition, Springer 2020. (book abbreviation: PDM). Library link: [Click here](#).

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2023

Brad Boehmke & Brandon M Greenwell, *Hands-on Machine Learning with R*, CRC 2019. (book abbreviation: HMLR). Library link: [Click here](#).

Brett Lantz, *Machine Learning with R*, 3rd edition, Packt Publishing 2019 (ISBN: 978-1788295864; book abbreviation: MLR). Amazon link: [Click here](#).

Other References: 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 [Canvas](#) website.

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 '23 Calendar: <https://registrar.camden.rutgers.edu/academic-calendar-2022-2023>

Grading: **Contributions to Final Grade**

- DataCamp courses	5%
- Take-home problems	25%
- Midterm Exams (2)	40%
- Final Exam or Project	25%
- 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.

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2023

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 2023

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)
Midterm Exam 1	Date: TBA
Topic 5	Naïve Bayes
Topic 6	Decision Trees & Ensemble methods using Random Forest
Topic 7	Logistic Regression Model
Topic 8	Linear Model Selection and Regularization
Midterm Exam 2	Date: TBA
Topic 9	Support Vector Machines
Topic 10	Association Rules (aka Market Basket Analysis)
Topic 11	Text Mining
Topic 12	Introduction to Neural Networks
Final Exam/Project	6:00-9:00 pm, Tuesday, May 9, 2023