

## OR/SYST 568: Applied Predictive Analytics (Spring 2020 - Tentative)

Instructor: Dr. Ran Ji  
Office: Nguyen Engineering Building, Room 2210  
Email: [rji2@gmu.edu](mailto:rji2@gmu.edu)  
Phone: (703) 993-5993  
Office Hours: Thursdays 4:30 – 6:30 PM or by appointment

**Place/Time:** Planetary Hall 212, Thursday 7:20 – 10:00 PM

**Recommended Prerequisites:** STAT 515 or Graduate Standing at the MSOR or MSSE programs. (Undergraduate engineering math: Calculus, probability theory, statistics, and some basic computer programming skills.)

**Course Description:** Introduces predictive analytics with applications in engineering, business, health care, marketing, and social economic areas. Topics include cross-sectional data processing, data visualization, correlation, linear and multiple regressions, classification and clustering, factor models, and predictive modeling performance analysis. Provides a foundation of basic theory and methodology with applied examples to analyze large engineering, social, and econometric data for predictive decision making. Hands-on experiments with R will be emphasized.

**Software:** R Software will be mainly used in this course. Class texts, lectures, and assignments provide substantial instructions and resources.

R and selected R packages constitute the primary software for this class. You can choose other software such as Python (but the coding work is not covered in class). R is the dominant software package for real world Predictive Analytics and will be used throughout other courses. R is an open-source software, RStudio is the recommended integrated development environment for using R. They can be downloaded at:

- R Software: <https://www.r-project.org>
  - RStudio Software: <https://www.rstudio.com/home/>
- To get familiar and started with R, you can visit and walk through:
- Joseph Adler: [R in a Netshell](#) (2<sup>nd</sup> edition)
  - W. N. Venables and D. M. Smith: [An Introduction to R](#)

**Textbooks:** **Suggest Textbooks and References:**

- Kuhn, Max, and Kjell Johnson. *Applied Predictive Modeling*, Springer, 2013.
  - Textbook Website: <http://appliedpredictivemodeling.com/>
  - E-Prints Download (Available via GMU Library & Springer): <https://link-springer-com.mutex.gmu.edu/content/pdf/10.1007%2F978-1-4614-6849-3.pdf>
- James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013.
  - Textbook Website: <http://www-bcf.usc.edu/~gareth/ISL/>
  - E-Prints Download (Available Online): <http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf>

- Rafael A. Irizarry. Introduction to Data Science: Data Analysis and Prediction Algorithms with R. CRC Press, 2020.  
- Textbook Website: <https://rafalab.github.io/dsbook/>
- Hastie, Trevor, Robert Tibshirani, and Jerome Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2009.
- Larose, Daniel T., and Chantal D. Larose. *Data Mining and Predictive Analytics* (2<sup>nd</sup> edition), John Wiley & Sons, 2015.
- Diez, David M., Christopher D. Barr and Mine Cetinkaya-Rundel. *OpenIntro Statistics* (3<sup>rd</sup> edition), Vol.12. CreateSpace, 2012.
- Hyndman, Rob J., and George Athanasopoulos. *Forecasting: Principles and Practice* (2<sup>nd</sup> edition). OTexts, 2018.

**Grading:** The course grade will be based on 4 individual assignments, 1 team assignment, an in-class midterm exam, a semester-long team project. Each grading component is described below.

Homework: 30%

Midterm Exam: 30%

Final Project: 40%

The +/- grades will be used.

The instructor reserves the rights to make minor modifications in the percentages related to the number of the assignments given.

**Final Projects:** Perform statistical analysis and predictive analytics on a practical problem. The team must define the problem subject to their expertise, preferences and interests, find the real or hypothetical data, propose and implement the applied predictive analytics modeling techniques, run the tests with comparisons and validations, derive practical insights and draw the final conclusions.

**Class Website:** **Blackboard:** <http://mymson.gmu.edu>

Click on the Courses tab in the green area (top right-of-center) and then on the OR-568 / SYST-568 (Spring 2020) link when the course list column appears.

**Syllabus:** Class syllabus, schedule, course overview information.

**Content:** Links to weekly modules with lectures, data files, R scripts and functions, and supplemental readings.

**Assignments:** Homework/Projects information, data and guidance. Solutions to assignments. Links to submit assignments/projects.

**Software:** Some resources and materials for R are available here.

**My Grades:** This is the place to check on your grades.

**Discussion Board:** One discussion board will be open to support communication among students.

**Course  
Schedule  
(Tentative):**

- **Week 1:** Introduction to Predictive Analytics and Review of Basic Statistics.
- **Week 2:** Data Pre-processing and PCA
- **Week 3:** Linear Regression Models I: OLS and Diagnostics
- **Week 4:** Linear Regression Models II: Cross-Validation, Ridge and Lasso
- **Week 5:** Linear Classification Model: Logistic Regression
- **Week 6:** Tree-based Models I: Classification and Regression Tree (CART)
- **Week 7:** Tree-based Models II: Random Forest and Boosting
- **Week 8:** Spring Break
- **Week 9:** In-Class Midterm Exam
- **Week 10:** Support Vector Machine (SVM)
- **Week 11:** Neural Network and Regression Splines
- **Week 12:** Naïve Bayes, KNN, Discriminant Analysis
- **Week 13:** Association Rules and Market Basket Analysis
- **Week 14:** Time-Series Forecasting
- **Week 15:** Final Project Presentations
- **Exam Week:** Final Project Report Due

---

## General Material

### Academic Integrity

Mason is an Honor Code university; please see University Catalog for a full description of the code and the honor committee process. The principle of academic integrity is taken very seriously and violations are treated gravely.

### Mason Email Accounts

Students must use their MasonLive email account to receive important University information, including the messages related to this class. See Mason Live (<http://masonlive.com>) for more information.

### Office of Disability Services

If you are a student with a disability and you need academic accommodations, please see me and contact the Office of Disability Services (ODS) at (703) 993-2474. All academic accommodations must be arranged through the ODS (<http://ods.gmu.edu>).

### Other Resources

**Writing Center:** Robinson Hall A114. Phone: (703) 993-1200. Webpage: <http://writingcenter.gmu.edu>

**University Libraries:** “Ask a Librarian”. Webpage: <http://library.gmu.edu/mudge/IM/IMRef.html>

**Counseling and Psychological Services (CAPS):** Phone (703) 993-2380. Webpage: <http://caps.gmu.edu>

**University Policies:** The University Catalog (<http://catalog.gmu.edu>) is the central resource for university policies affecting student, faculty, and staff conduct in university academic affairs. Other policies are available at University Policy (<http://universitypolicy.gmu.edu>). All members of the university community are responsible for knowing and following established policies.