MIS 6357: Advanced Business Analytics with R

Fridays, 4:00pm - 6:45pm, JSOM 1.217

Instructor: Brian Lois
bxl171630@utdallas.edu
Office Location: To be determined
Office Hours: After class

Course Description: This course is based on the open-source R software. Topics include data manipulation, imputation, variable selection, as well as advanced analytic methods
Prerequisite: MIS 6356
Credit Hours: 3

Text: *Applied Predictive Modeling*
Authors: Max Kuhn and Kjell Johnson; **ISBN:** 978-1-4614-6848-6

Grade Distribution:

Homework	30%
Quizzes	10%
Exam 1	20%
Exam 2	20%
Final Project	20%

Course Policies:

- **General**

  - Quizzes and exams are closed book, closed notes.
  - **No makeup quizzes or exams will be given.**

- **Assignments**

  - Students are expected to work independently. **Offering** and **accepting** solutions from others is an act of **academic dishonesty**, which is a serious offense and **all involved parties will be penalized according to the Academic Honesty Policy**. Discussion amongst students is encouraged, but no written notes may be retained from those discussions.
  - You are allowed to search the internet, but my **never** copy any code. Your solution should be written while not looking at others' code.
  - Posting homework related questions to websites such as stackoverflow is prohibited.
  - **No late assignments will be accepted**.

- **Attendance and Absences**

  - Attendance is expected.
  - You are responsible for any announcements made in class.

Tentative Course Outline:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments.

Week	Content
Week 1	- Introduction to R

Week 2	- Introduction to modeling
	- Data pre-processing
	  – Principal Components Analysis
	  – Missing value imputation
	- Reading assignment: Ch. 2-3

Week 3	- Over fitting and tuning
	  – K-fold cross validation
	  – Bootstrapping
	  – Grid searching
	- Reading assignment: Ch. 4

Week 4	- Measuring model performance
	  – RMSE, R^2
	  – Bias variance trade-off
	  – Sensitivity, Specificity, ROC curves and AUC
	- Reading assignment: Ch. 5 and 11

Weeks 5-6	- Linear models
	  – Linear regression, logistic regression
	  – Penalized regression: ridge regression, lasso
	  – Partial least squares
	- Reading assignment: Ch. 6 and 12

Weeks 7-9	- Non-linear models
	  – Neural networks
	  – Multivariate Adaptive Regression Splines
	  – Support vector machines
	  – K-nearest neighbor
	- Reading assignment: Ch. 7 and 13

Weeks 10-11	- Tree based models
	  – Random forests
	  – Gradient Boosting Machines (XGBoost)
	- Reading assignment: Ch. 8 and 14

Week 12	- Class imbalance - Reading assignment: Ch. 16
Week 13	- Feature Importance and Selection – Recursive feature elimination - Reading assignment: Chs. 18 and 19