

# 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	<ul style="list-style-type: none"><li>• Introduction to R</li></ul>
Week 2	<ul style="list-style-type: none"><li>• Introduction to modeling</li><li>• Data pre-processing<ul style="list-style-type: none"><li>– Principal Components Analysis</li><li>– Missing value imputation</li></ul></li><li>• Reading assignment: Ch. 2-3</li></ul>
Week 3	<ul style="list-style-type: none"><li>• Over fitting and tuning<ul style="list-style-type: none"><li>– K-fold cross validation</li><li>– Bootstrapping</li><li>– Grid searching</li></ul></li><li>• Reading assignment: Ch. 4</li></ul>
Week 4	<ul style="list-style-type: none"><li>• Measuring model performance<ul style="list-style-type: none"><li>– RMSE, <math>R^2</math></li><li>– Bias variance trade-off</li><li>– Sensitivity, Specificity, ROC curves and AUC</li></ul></li><li>• Reading assignment: Ch. 5 and 11</li></ul>
Weeks 5-6	<ul style="list-style-type: none"><li>• Linear models<ul style="list-style-type: none"><li>– Linear regression, logistic regression</li><li>– Penalized regression: ridge regression, lasso</li><li>– Partial least squares</li></ul></li><li>• Reading assignment: Ch. 6 and 12</li></ul>
Weeks 7-9	<ul style="list-style-type: none"><li>• Non-linear models<ul style="list-style-type: none"><li>– Neural networks</li><li>– Multivariate Adaptive Regression Splines</li><li>– Support vector machines</li><li>– K-nearest neighbor</li></ul></li><li>• Reading assignment: Ch. 7 and 13</li></ul>
Weeks 10-11	<ul style="list-style-type: none"><li>• Tree based models<ul style="list-style-type: none"><li>– Random forests</li><li>– Gradient Boosting Machines (XGBoost)</li></ul></li><li>• Reading assignment: Ch. 8 and 14</li></ul>
Week 12	<ul style="list-style-type: none"><li>• Class imbalance</li><li>• Reading assignment: Ch. 16</li></ul>
Week 13	<ul style="list-style-type: none"><li>• Feature Importance and Selection<ul style="list-style-type: none"><li>– Recursive feature elimination</li></ul></li><li>• Reading assignment: Chs. 18 and 19</li></ul>