

## STA 6704 Data Mining II (Spring 2016)

### [STA 6704] – [Data Mining Methodology II] [Fall] – [Statistical Computing] – [3 Credit Hours]

<b>INSTRUCTOR:</b>	Morgan C. Wang
<b>PHONE:</b>	407-823-2818
<b>OFFICE LOCATION:</b>	TC II 203
<b>OFFICE HOURS:</b>	Monday and Wednesday 1:30 AM to 4:00 PM
<b>E-MAIL:</b>	<a href="mailto:Chung-Ching.Wang@ucf.edu">Chung-Ching.Wang@ucf.edu</a>
<b>WITHDRAW DATELINE:</b>	March 23, 2016
<b>HOLIDAYS:</b>	January 18 and March 7 to March 12
<b>SPECIAL NOTES:</b>	<b>Students who are not officially registered in the class will not have exams graded or returned.</b>

---

#### LEARNING OBJECTIVES:

At the end of the course, students will be able to:

- Learning Objective 1: Use SAS/Enterprise Miner/R software effectively to perform data mining exercises
- Learning Objective 2: Use clustering techniques to perform customer segmentation and clustering analysis
- Learning Objective 3: Use association analysis to perform market basket analysis and understand variable relationship
- Learning Objective 4: Use neural network techniques to build predict models and perform clustering
- Learning Objective 5: Use text data along with numerical data to build classification models and predictive model
- Learning Objective 6: Use text data along with numerical data to perform cluster analysis

#### COURSE DESCRIPTION:

Statistical component of unsupervised mining techniques for mining big data. Topics include association analysis, clustering analysis, text clustering, time series clustering, discriminant analysis, and factor analysis. Prerequisite: STA 5703, Basic Statistical knowledge, Programming Language/SAS, and Enterprise Miner/R

## STA 6704 Data Mining II (Spring 2016)

### PURPOSE OF THE COURSE:

Statistical learning is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of observational data in order for the data owner to discover meaningful patterns and models. Analyst can use un-supervise learning techniques such as association analysis and clustering analysis to enhance data quality. Consequently, the analysts can use the enhanced data to build better predictive models. Analyst can also use unstructured data such as text data, network data, and time series data to enhance the quality of the data. By combining both structure data and unstructured the quality of predictive model can be significantly enhanced. This course will focus on understanding unsupervised learning techniques and unstructured data. Then by combining the data elements created using unsupervised learning and unstructured data with traditional structure data to solve the real world problems more effectively. The high quality model produced can then be used to support decision-making and to identify new business opportunities.

### REQUIRED MATERIALS:

#### Required Text

- Lecture Notes from the instructor
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) by Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications, Elsevier by Gary D. Miller, Dursun Dalen, John F. Elder, Andrew Fast, Thomas Hill, and Robert A. Nisbet.

#### Supplemental Materials

- Selected articles by the instructor

### ATTENDANCE POLICY:

There are pop up quizzes every week and students who miss pop up quiz without advanced permission from the instructor will get zero score of that quiz. Students who miss two or more quizzes without adequate excuse and advance permission from the instructor will get 25 additional points deduction from their final grade. It is your responsibility to attend all classes and to notify instructor all absences in advance and provide instructor the related documents.

### MAKE-UP EXAM POLICY:

Make-up exams will be allowed only in extreme instances and with advanced permission of the instructor. It is the student's responsibility to work with faculty to notify them of an excused absence (e.g. work, illness) and to coordinate a make-up exam. EDC staff are available to administer and proctor make-up exams during the EDC's regular operating hours.

## STA 6704 Data Mining II (Spring 2016)

### ACADEMIC INTEGRITY:

All students are expected to abide by the University's Code of Student Conduct.

Plagiarism and Cheating of any kind on an examination, quiz, or assignment will result at least in an "F" for that assignment (and may, depending on the severity of the case, lead to an "F" for the entire course) and may be subject to appropriate referral to the Office of Student Conduct for further action. See the [UCF Golden Rule](#) for further information.

### GRADING SCALE:

LETTER GRADE	PERCENTAGE
A	92.5-100 %
A-	90.0-92.4 %
B+	87.0-89.9%
B	83.0-86.9%
B-	80.0-82.9%
C+	77.0-79.9%
C	73.0-76.9%
C-	70.0-72.9%
D+	67.0-69.9%
D	63.0-66.9%
D-	60.0-62.9%
F	Less than 59.9%

### GRADING PROCEDURES:

All quizzes in this semester worth 30% of the final grade. Each quiz worth approximately 30 points. All assignments in this semester worth 20% of the final grade. Each assignment worth approximately 30 points.

ASSIGNMENT	% OF GRADE
Homework Assignments	20%
Exam 1 / Midterm Project (March 14 and 16, 2015 6:00 PM to 7:15 PM)	25%
Weekly Pop Quizzes	30%

## STA 6704 Data Mining II (Spring 2016)

Exam 2 / Final Project (May 2, 2016 4:00 PM to 6:50 PM)	25%
Total	100%

### ASSIGNMENT DESCRIPTIONS:

Pop up quiz will be given at least once each week. After the completion of each lecture, the lecture material will be covered in the weekly pop quiz. Exam I is individual data mining project sponsored by Everbank. Exam II is SAS Shootout competition, a team big data analytical project. The final project includes weekly meeting, progress report, presentation, analysis, and final report. All student must join one team and complete the final report to get grade. Assignments will be given six times during the semester after week #3. All assignments are analytical projects using supervised and/or un-supervised learning techniques.

### COURSE OUTLINE:

CLASS DATE	TOPICS	READING(S) & ASSIGNMENT(S)	DUE DATE
<b>Week #1</b>	Supervised Learning III – Neural Network Fundamental (Review)	Lecture Note # 4, 5, 6	
<b>Week #2</b>	Supervised Learning III – Neural Network Modeling (Review)	Lecture Note # 7, 8, 9	
<b>Week #3</b>	Lecture #1: Using Enterprise Miner/R	Lecture Note # 1	
<b>Week #4</b>	Lecture #2: Clustering and Customer Segmentation	Lecture Note # 2	
<b>Week #5</b>	Lecture #2: Clustering and Customer Segmentation	Lecture Note # 2	
<b>Week #6</b>	Lecture #3: Association Analysis	Lecture Note # 3	
<b>Week #7</b>	Lecture #3: Association Analysis	Lecture Note # 3	
<b>Week #8</b>	Everbank Competition Winner Presentation and EXAM I		
<b>Week #9</b>	SAS Shootout First Round Presentation		
<b>Week #10</b>	Lecture #10 Text Miner Nodes Overview	Lecture Note # 10	
<b>Week #11</b>	Lecture #11 Text Data Preparation using Text Miner	Lecture Note # 11	
<b>Week #12</b>	Lecture #12: Model Enhancement using Textual Data (Clustering)	Lecture Note # 12	
<b>Week #13</b>	Lecture #13: Model Enhancement using Textual Data (Predictive Modeling)	Lecture Note # 13	
<b>Week #14</b>	Lecture # 14: Analysis Time Series Data	Lecture Note # 14	
<b>Week #15</b>	Case Studies on Big Data Applications		

**Note: Syllabus subject to change based on needs of students, University, and instructor. All material covered in class, regardless of whether material is listed, is fair to be tested.**

## STA 6704 Data Mining II (Spring 2016)

### DISABILITY STATEMENT (OPTIONAL)

The University of Central Florida is committed to providing reasonable accommodations for all persons with disabilities. This syllabus is available in alternate formats upon request. Students with disabilities who need accommodations in this course must contact the professor at the beginning of the semester to discuss needed accommodations. No accommodations will be provided until the student has met with the professor to request accommodations. Students who need accommodations must be registered with [Student Disability Services](#), Ferrell Commons, 7F, Room 185, phone (407) 823-2371, TTY/TDD only phone (407) 823-2116, before requesting accommodations from the professor.

### COPYRIGHT (OPTIONAL)

This course may contain copyright protected materials such as audio or video clips, images, text materials, etc. These items are being used with regard to the Fair Use doctrine in order to enhance the learning environment. Please do not copy, duplicate, download or distribute these items. The use of these materials is strictly reserved for this online classroom environment and your use only. All copyright materials are credited to the copyright holder.

**\*NOTE:** For additional sample syllabi information, including copy ready syllabi clauses related to using Webcourses, Turnit in, etc., please visit UCF's Faculty Center for Teaching & Learning: <http://www.fctl.ucf.edu/TeachingAndLearningResources/CourseDesign/Syllabus/statements.php#ethics>