

# Big Data Analytics

CSCE 587 Big Data Analytics  
Fall 2016  
Swearingen 2A31  
TTh 10:05-11:20am

John Rose  
[rose@cse.sc.edu/7-2405](mailto:rose@cse.sc.edu/7-2405)  
Office: Swearingen 3A67  
Office hours: TW 3:30-5pm  
TA: No S. Luck

## Course Description

CSCE 587 – Special Topics: Big Data Analytics (3) This course covers foundational techniques and tools required for data science and big data analytics. The course focuses on concepts, principles, and techniques applicable to any technology environment and industry and establishes a baseline that can be enhanced by further formal training and additional real-world experience.

## Course Overview

This course introduces the student to concepts of big data management, data mining techniques and the underlying statistics that support big data analytics. Since this is a 500-level course, relevant 500 level courses such as Database (CSCE 520) and Statistics for Engineers (STAT 509) cannot be listed as prerequisites. Consequently, lecture time will be devoted to addresses necessary topics from these courses. In this course we will use the programming language R as the primary tool for analysis.

## Learning Outcomes

By the end of the course the student will be able to:

1. deploy a structured lifecycle approach to data science and big data analytics projects
2. select visualization techniques and tools to analyze big data and create statistical models
3. use tools such as R and RStudio, and MapReduce/Hadoop.

## Required Texts, Other Materials, Suggested Readings

This course does not have a required text. However, ad hoc readings from the field will be assigned. In addition, material from “Data Science and Big Data Analytics Student Guide” distributed by EMC Education Services will be provided to the students.

## Course Delivery Structure

The course will be delivered in a computer lab with 50% of the time devoted to lectures and the other 50% devoted to hands-on lab exercises.

## Course Requirements

Readings: Students will read lecture material assigned for each class prior to the class.

Homework: Students will complete assignments demonstrating mastery of material. These will be due at the beginning of class.

## Course Outline/Schedule

Lecture 1: Introduction to Big Data Analytics

Lecture 2: DBMS Overview

Lecture 3: Introduction to R and RStudio

Lecture 4: Basic analysis in R  
Lecture 5: Intermediate R  
Lecture 6: Intermediate analysis in R  
Lecture 7: Visualization and Data Exploration  
Lecture 8: K-means Clustering.  
Lecture 9: Independent Sample Tests  
Lecture 10: Basic Association Analysis  
Lecture 11: Association Rule Speedup  
Lecture 12: Linear regression part 1  
Lecture 13: Linear regression part 2  
Lecture 14: Logistic regression  
Lecture 15: Naïve Bayes  
Lecture 16: Decision trees part 1  
Lecture 17: Decision trees part 2  
Lecture 18: Review for Midterm Exam  
Lecture 19: Midterm Exam  
Lecture 20: Introduction to Hadoop and HDFS  
Lecture 21: Using R with Hadoop  
Lecture 22: First R/Hadoop program  
Lecture 23: Intermediate R/Hadoop programming  
Lecture 24: Pig, Hive, and HBase  
Lecture 25: Discussion of rmr2 Project  
Lecture 26: Support Vector Machines Part 1  
Lecture 27: Support Vector Machines Part 2  
Lecture 28: Review for Final Exam

### **Assignments**

*Readings:* Students will read lecture material assigned for each class prior to the class. Readings will be assigned at the end of the preceding class.

*Homework:* Students will complete assignments demonstrating mastery of material. These will be due at the beginning of class. Graded written evaluations will be returned one week after submission.

HW 1: R Homework assignment

HW 2: K-means homework assignment

HW 3: Association Analysis homework assignment

HW 4: Linear and logistic regression homework assignment

HW 5: Naïve Bayes homework assignment

HW 6: Decision tree homework assignment.

HW 7: rmr2 Project

Midterm exam: Covers lectures 1 – 17: In-class exam as well as take-home applied-exam

Final exam: Covers entire semester: In-class exam as well as take-home applied-exam

**Grading Scheme**

Final grade:  $90 \leq A$ ,  $87 \leq B+ < 90$ ,  $80 \leq B < 87$ ,  $77 \leq C+ < 80$ ,  $65 \leq C < 77$ ,  $60 \leq D+ < 65$ ,  $50 \leq D < 60$ ,  $F < 50$

Grades will be calculated from homework (30%), in-class labs (20%), midterm (20%), and Final Exam (30%).

**Difference between Undergraduate and Graduate Work:**

Graduate students are assigned additional problems in both homework and exams.

**Course Policies**

*Attendance:* Attendance is mandatory. Students will be expected to have read the material for each lecture prior to the lecture and to be able to actively participate in discussions during class.

*Tardiness, late assignments:* homework is due at the beginning of class. Late assignments will be charged 20% per day.

*Violations of academic honesty:* Assignments and examination work are expected to be the sole effort of the student submitting the work. Students are expected to follow the University of South Carolina Honor Code and should expect that every instance of a suspected violation will be reported. Students found responsible for violations of the Code will be subject to academic penalty under the Code in addition to whatever disciplinary sanctions are applied.

*Policy on disabilities or special needs:* Any student with a documented disability should contact the Office of Student Disability Services at 803-777-6142 to make arrangements for appropriate accommodations.