

---

# TENNESSEE TECH UNIVERSITY

## DEPARTMENT OF COMPUTER SCIENCE

### CSC-3220: FUNDAMENTALS OF DATA SCIENCE

TR, 9:30-10:50AM, BR 406, 3 CREDIT HOURS, SPRING 2017

#### INSTRUCTOR INFORMATION

**Instructor's Name:** William (Bill) Eberle

**Telephone Number:** 931-372-3278

**Email:** [weberle@tntech.edu](mailto:weberle@tntech.edu)

**Office:** Bruner Hall, Rom 413

**Office hours:** Mondays 9:00-10:30 and Tuesdays 1:00-2:30

#### COURSE INFORMATION

##### PREREQUISITES

CSC 2400, MATH 2010, MATH 3470 or MATH 3070

##### COURSE DESCRIPTION

The purpose of this course is to introduce students, who have already achieved a solid algorithmic, programming and mathematical background, to the tools and techniques for developing data science applications. This course serves as an introduction to the basics of data science including programming for data management, data manipulation, data analytics, and data visualization. Students will be introduced to various machine learning algorithms, and learn to formulate context-relevant questions and hypothesis to drive scientific research and understand statistical inference. Students will be introduced to the R programming language. The foundation is laid for big data applications ranging from fraud detection to healthcare informatics.

##### COURSE OBJECTIVES

- ✓ Provide students with fundamental knowledge and training for analyzing and solving data science problems in professional practice and for undertaking graduate studies.
- ✓ Equip students to apply industry-standard, data science principles across various types of data science problems.
- ✓ Develop written and oral communication skills in students that are necessary to function effectively in the profession and society.
- ✓ Provide students with a broad education that:
  - develops an appreciation and understanding of current issues in data science and their impact on social and global issues,
  - instills an understanding of professional and ethical responsibilities necessary for the complex modern work environment, and
  - develops the desire and abilities needed for lifelong learning.

## STUDENT LEARNING OUTCOMES

- ✓ Ability to apply knowledge of data science in the field of computer science.
- ✓ Ability to apply fundamental and advanced concepts of data science.
- ✓ Ability to apply critical thinking skills to analyze and evaluate potential data science solutions.
- ✓ Ability to convey technical data science material through written materials and through interacting with an audience in an oral presentation.
- ✓ Understanding professional, legal, and ethical responsibilities associated with the field of data science.
- ✓ Experience with how data science knowledge is applied in advanced areas of the field.

## MAJOR TEACHING METHODS

Lecture, discussion, and in-class assignments.

## SPECIAL INSTRUCTIONAL PLATFORM/MATERIALS

(Required) Laptops with RStudio/R installed.

## TOPICS TO BE COVERED:

1. DATA SCIENCE PROCESS
2. INTRODUCTION TO R
3. DATA ACQUISITION AND CLEANING
4. WORKING WITH DATES AND TIMES
5. TEXT MINING AND NATURAL LANGUAGE PROCESSING
6. DATA VISUALIZATION
7. EXPLORATORY ANALYSIS
8. MANAGING DATA
9. STATISTICAL INFERENCE
10. LOGISTIC REGRESSION
11. SOCIAL NETWORKS
12. MAPREDUCE AND HADOOP
13. PRESENTING RESULTS

## TEXTS AND REFERENCES:

**Required:** Notes and various resources will be provided.

### References:

Matloff, The Art of R Programming (free PDF download)  
Zumel & Mount, Practical Data Science with R (free PDF download)  
Wickham & Golemund, R for Data Science (free online)

## GRADING AND EVALUATION PROCEDURES:

LABS	20%
HOMEWORK	20%
TEAM PROJECT	30%
EXAMS	30%

## GRADING SCALE

Letter Grade	Grade Range
A	90-100
B	80-89
C	70-79
D	60-69
F	59 and below

## COURSE POLICIES

### STUDENT ACADEMIC MISCONDUCT POLICY

Maintaining high standards of academic integrity in every class at Tennessee Tech is critical to the reputation of Tennessee Tech, its students, alumni, and the employers of Tennessee Tech graduates. The Student Academic Misconduct Policy describes the definitions of academic misconduct and policies and procedures for addressing Academic Misconduct at Tennessee Tech. For details, view the Tennessee Tech's Policy 217 – [Student Academic Misconduct at Policy Central](#).

### CLASS PARTICIPATION

The course is highly participatory in regards to learning activities performed in class. As such, students are expected to attend and actively engage with the instructor and fellow students in the class.

### ASSIGNMENTS AND RELATED POLICY

The assignments are intended to provide students with the opportunity to apply what they are learning in this course. The duration of assignments will vary. **No late assignments will be accepted.**

### Class Plan by Weeks or Days:

Week	Dates	Topic	Project Deliverables
1	01/17–01/19	Data Science Process	
2	01/24-01/26	Introduction to "R"	
3	01/31-02/02	Statistical Inference	
4	02/07-02/09	Exploratory Data Analysis	
5	02/14-02/16	Algorithms and Linear Regression	
6	02/21-02/23	Logistic Regression	Teams and Project Preferences
7	02/28-03/02	Data Acquisition Data Cleaning	Project Proposals Research Day Abstracts
SPRING BREAK			

8	03/14-03/16	Midterm Exam Text Mining	
9	03/21-03/23	Natural Language Processing Data Visualization	
10	03/28-03/30	Data Visualization	
11	04/04-04/06	Model Evaluation	Research Day Poster
12	04/11-04/13	Causality	
13	04/18-04/20	MapReduce and Hadoop Social Networks	
14	04/25-04/27	Presenting Results <i>Team Project Presentations</i>	Project Report Team Presentation
15	05/01-05/05	Final Exam	

#### DISABILITY ACCOMMODATION

Students with a disability requiring accommodations should contact the Office of Disability Services (ODS). An Accommodation Request (AR) should be completed as soon as possible, preferably by the end of the first week of the course. The ODS is located in the Roaden University Center, Room 112; phone 372-6119. For details, view the Tennessee Tech's Policy 340 – [Services for Students with Disabilities at Policy Central](#).