STAT 1650: Introduction to Data Science

2021 Winter Session	
Total Class Sessions: 25	**Instructor: Staff**
Class Sessions Per Week: 6	**Classroom: TBA**
Total Weeks: 4	**Office Hours: TBA**
Class Session Length (Minutes): 145	**Language: English**
Credit Hours: 4	

Course Description:

Data science is a field that combines mathematics, computer science and statistics, with the goal of answering questions and discovering new information from data. Data has become an increasingly popular and reliable tool, which more and more organizations adopt to support their decision and build data-intensive products and services. Therefore, data science is an important field to support those functions and develop those skills.

This course will cover essential elements of data science including data collection and management, summarizing and visualizing data, utilizing basic ideas of statistical inference, and machine learning. Students will gain hands-on experience through using statistical software.

Course Objectives:

Upon completion of the course, you will be able to:

- Identify and describe the methods and techniques commonly used in data science.
- Demonstrate proficiency with the methods and techniques for obtaining, organizing, exploring, and analyzing data.
- Utilize statistical concepts of data analysis, data collection, modeling, and inference.
- Develop basic skill of Python programming and its application in data science.
- Apply modern data science methods to one or more domains of application (e.g. business analytics, finance, biotechnology, and public health)

Course Materials:

Recommended Books:

Data Science for business, F. Provost, T Fawcett, 2013

Python for Data Analysis, Wes McKinney, O'Reilly , 2012

Assessment Summary:

Class Participation (5%)

Active participation and interaction with your peers is an important learning activity. Please refer to the relevant instruction document for your class mode (face-to-face, virtual, or web) to learn more about the participation requirements and scoring.

Quizzes (10%)

This assessment includes a series of weekly online quizzes that you can perform in the comfort of your own home. There are 5 quizzes based on. Each quiz consists of 3 questions. The quizzes contain both multiple-choice and calculation questions. For each quiz you will have one hour. The quiz will be available from Thursday of the previous week to Sunday of the week it is due. There are 5 quizzes in total. Your best 3 quizzes will count towards the final mark.

Group Assignment (20%)

The general purpose of a case study is to describe an individual situation (case), e.g. person, business, organization, or institution, in detail and identify the key issues of the case (your assignment question should tell you what to focus on). You need to analyze the case using relevant theoretical concepts from your unit or discipline and recommend a course of action for that particular case (particularly for problem-solving case studies).

Midterm and Final Exam (25 + 40 %)

There is a midterm and a final exam. Both exams are open book and you are free to bring a calculator to the exam. The exams will consist on Numerical, Theoretical and conceptual (applied) questions. The exam dates are provided in the Course Schedule attached to the end of this syllabus. There are no makeup exams.If you miss the midterm exam, regardless of the reason, your final score will be used in its place. If you miss the final exam, you will automatically fail the course. The only excuse for missing the final exam is a medical emergency, the validity of which will be verified by the University. The best way to prepare for the exams is to come to class and participate, work all of the cases, and complete the homework assignments.

Class Participation	5%
Quizzes	10%
Group Assignment	20%
Midterm	25%
Final Exam	40%
Total	100%

Grading Scale (percentage):

A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
98-100	93-97	90-92	88-89	83-87	80-82	78-79	73-77	70-72	68-69	63-67	60-62	<60

Academic Integrity:

Students are encouraged to study together, and to discuss lecture topics with one another, but all other work should be completed independently.

Students are expected to adhere to the standards of academic honesty and integrity that are

described in the Chengdu University of Technology's *Academic Conduct Code.* Any work suspected of violating the standards of the *Academic Conduct Code* will be reported to the Dean's Office. Penalties for violating the *Academic Conduct Code* may include dismissal from the program. All students have an individual responsibility to know and understand the provisions of the *Academic Conduct Code*.

Special Needs or Assistance:

Please contact the Administrative Office immediately if you have a learning disability, a medical issue, or any other type of problem that prevents professors from seeing you have learned the course material. Our goal is to help you learn, not to penalize you for issues which mask your learning.

Course Schedule:

Tentative Class Schedule

The instructor reserves the right to make changes to this syllabus based on course needs

Week No.	Topics	Activities	Assignment
1	**Course Introduction and Overview:** • Syllabus • Overview of the course **Foundations of Data Science** • Introduction to Big Data, Data Science, and Predictive Analytics • Data science Concepts and Application Areas • Introduction to Python Programming • Comparison of Python, R and Matlab usage in data science • Fundamentals of Machine Learning • Data Exploration, Data visualization and Analytics in Python • Hands-On Lab: Data Exploration, Data visualization and Analytics in Python **Basic Statistics** • Random Variables, Sampling • Distributions and Statistical Measures • Hypothesis Testing • Hands-On Lab: Case Study and Python	Learning targets, Dos and Do-Not's. Briefing on Project/ assignments/ Discussion /Lecture	Quiz 1 Assignment 1

	programming		
2	**Overview of Linear Algebra** • Linear algebra and matrix computations • Functions, derivatives, convexity **Optimization** • Unconstrained non-linear optimization algorithms • Overview of constrained optimization algorithms **Modeling Techniques, Regression** • Mathematical modeling process • Linear regression • Logistic regression • Regression case studies • Hands-On Lab: Building a Regression Model	Lecture/Presentation/Discussions After Class Discussion questions	Quiz 2 Quiz 3 Assignment 2
3	**Classification Algorithms** • Introduction to Predictive Modeling • Decision Tree Learning • Logistic Regression • Naïve Bayes • Hands-On Lab: Building a Classifier **Regression Algorithms** • Linear Regression • Regularized Regression Models • Hands-On Lab: Building a Regression Model **Unsupervised Learning** • K-Means Clustering • Hands-On Lab: Using K- Means Clustering	Lecture/Presentation/Discussions After Class Discussion questions	Midterm Exam Assignment 3 Quiz 4

4	**Recommender Systems** • Text Analytics • Content-Based and Collaborative Filtering • Evaluation of Recommendation Systems. DCG • Hands-On Lab: Analyzing a Document Collection **Ensemble Methods** • Bootstrapping, Bagging, and Boosting • Random Forests • Building a Random Forest Classifier • Calculating Probabilities with Binomial Distribution, Sampling with and without Replacement • Hands-On Lab: Building a Random Forest Classifier **Simulation modeling** • Random number generation • Monte Carlo simulations • Simulation case studies	Lecture/Presentation/Discussions After Class Discussion questions	Quiz 5 Assignment 4 Final Exam (Cumulative, TBA)