



The UNIVERSITY of OKLAHOMA
Gallogly College of Engineering
Data Science and Analytics

The following details the core courses each student is required to take:

- Page 2. DSA 5005 - Computing Structures
- Page 3. DSA 4513 - Database Management Systems
- Page 4. DSA 4413 - Algorithm Analysis
- Page 5. DSA 5013 - Statistical Analysis and System Design
- Page 6. DSA 5103 - Intelligent Data Analysis
- Page 7. DSA 5113 - Advanced Analytics & Metaheuristics
- Page 8. DSA 5900 - Internship or Practicum

DSA 5005 – Computing Structures

INSTRUCTORS: Sudarshan Dhall and Sridhar Radhakrishnan

PREREQUISITE: Permission of graduate liaison

TEXT: Data Structures Featuring C++: A Programmer’s Perspective by Radhakrishnan, S., Wise L., and Sekharan, N. 2013 and Applied Discrete Mathematics by Doerr, A. and Lefebvre, K. 2013 (this content is available for free on the Janux platform)

MATERIALS: In order to get the most out of this course, you will need to make sure that you have the following hardware, software or web accounts:

- | | |
|--|--|
| <ul style="list-style-type: none">• Webcam• Calculator• Wolfram CDF Player• Access to scanner or scanning App to PDF• C++ Compiler | <ul style="list-style-type: none">• B Virtual Proctoring account• Google + account for Google Hangout• Adobe Acrobat PDF reader (latest version)• Updated browser |
|--|--|

COURSE DESCRIPTION: This course addresses discrete mathematics, object-oriented programming in C++, and data structures in C++. As part of the discrete mathematics students will be introduced to combinatorics, logic, relations, functions, computational complexity, automata, and graph theory. Students will be introduced to the fundamentals of object-oriented programming and learn to design, build, and analyze data structures using object-oriented principles and techniques.

COURSE OVERVIEW: This course is organized by week and is divided into two main sections: (1) Data Structures and (2) Discrete Mathematics. In addition to watching screencasts, reading texts and taking notes, you will have weekly homework assignments (DM), three programming projects (DS), a mid-term and a final exam.

COURSE GOALS:

1. Object-Oriented Programming: Design, build and analyze data structures using object-oriented principles.
2. Object-Oriented Programming in C++ and Data Structures in C++.
3. Combinatorics, Logic, Relations, Functions, Computational Complexity, Automata and Graph Theory.

DSA 4513 - Database Management

INSTRUCTOR: Le Gruenwald

PREREQUISITE: CS 2413 or CS 5005, and CS 2813 or CS 4005 or MATH 2513

COURSE DESCRIPTION: This course concentrates on the concepts and structures necessary to design and implement a database system. It is oriented towards the general principles, their applications, and the relevant theoretical foundation. It does not devote any significant time to discussion of specific commercial packages.

TEXT: Database System Concepts, 6th Edition, by Abraham Silberschatz, Henry Korth, and S. Sudarshan, and A Gift of Fire, 4th Edition, by Sara Baase

COURSE TOPICS:

1. Introduction to Database Management Systems: Teach the components, purposes and evolution of database management systems to students who know discrete math and data structures.
2. Entity-Relationship Model: Teach the design process and concepts necessary to design a database using the Entity-Relationship model to students who know the functions of database management systems.
3. Relational Model & SQL: Teach the relation data model and the SQL language to students who know the functions of database management systems and how to design a database using the ER model.
4. File Organization, Indexing & Hashing: Learn different file organizations and how to choose an appropriate file organization for a database table to students who know the function of database management systems, the relational model, and the SQL language and data structures.
5. Formal Relational Query Languages: Teach the formal query languages of the relational data model to students who know the functions of database management systems, the relational model and the SQL language.
6. Functional Dependencies & Normalization: Teach the concepts of functional dependencies and normalization and their use in database design to students who know the functions of database management systems, the relational data model, the ER model and the SQL query language.

COURSE GOALS:

1. An understanding of professional, ethical, legal, security and social issues and responsibilities.
2. An ability to analyze the local and global impact of computing on individuals, organizations, and society.

DSA 4413 – Algorithm Analysis

INSTRUCTOR: S. Lakshmivarahan

PREREQUISITE: CS 2413 or CS 5005, and CS 2813 or CS 4005 or MATH 2513

COURSE DESCRIPTION: Design and analysis of algorithms and measurement of their complexity.

TEXT: Introduction to Algorithms by Corman, Leiserson and Rivest. McGraw Hill, Second/Third Edition

COURSE TOPICS:

1. Algorithm Analysis.
2. Structure of Algorithms: Recall various ways in which different algorithms behave
3. Comparison-based Problems: Recognize the tools of analysis. Describe the strategy of Divide and Conquer.
4. Upper and Lower Bound: Find the upper and/or lower bounds of a variety of sets or subsets.
5. Sorting: Using quicksort to place the elements of an array in order. Use heap sort to divide input into a sorted and unsorted regions and iteratively shrink the unsorted region by extracting the largest element and moving it to the sorted region.
6. Optimization Problems: Recognize and use two strategies for algorithm design: dynamic programming and greedy strategies.
7. Dynamic Programming: Implement Dynamic programming. Perform Matrix Chain Product
8. Strassen's Method: Implement Strassen's Algorithm for Matrix Multiplication. Obtain the complexity for Strassen's algorithm.
9. Graph Algorithms: Solve Graph Traversal Problems using DFS and BFS Techniques. Obtain Minimum Cost Graph using algorithms such as Minimum Spanning Tree, Prim's Algorithm, and Kruskal Algorithm.
10. Randomized Algorithms: Solve Problems using Randomized Algorithms. Apply Monte Carlo Method to solve problems using random number. Apply Las Vegas Algorithm to solve using random number.

COURSE GOALS:

1. An ability to acquire knowledge of computing and mathematics appropriate to the discipline.
2. An Ability to apply knowledge of computing and mathematics appropriate to the discipline.
3. An Ability to analyze a problem, and identify and define the computing requirements appropriate to its solution.

DSA 5013 – Statistical Analysis and System Design

INSTRUCTOR: Kash Barker and Charles Nicholson

PREREQUISITE: Graduate standing

COURSE DESCRIPTION: Introduction to probability, expectation, discrete and continuous distributions, sampling and descriptive statistics, parameter estimation, and statistical tests to aid decision making. The student will learn analysis techniques for verification of systems parameters.

TEXT: Statistics for Engineering and the Sciences. 6th edition, Mendenhall, W. and T. Sincich

COURSE TOPICS:

1. Descriptive Statistics.
2. Define events and use common rules of probability to calculate likelihood of events.
3. Discrete Random Variables: Define and interpret the meaning of a random variable.
4. Introduction to Continuous Random Variables: Understand the characteristics of a continuous random variable and its probability distribution.
5. Sampling Distributions: Understand that sample statistics are themselves random variables that follow a probability distribution.
6. Statistical Intervals for a Single Population.
7. Test of Hypotheses for a Single Population.
8. Statistical Inference for Two Populations.
9. Nonparametric Statistics: Differentiate between parametric and nonmethods and when each should be used.
10. Fitting Probability Distributions.
11. Implementing Statistical Tests with R Software: Install R on your computer. Read data from spreadsheets. Carry out various analyses from this course, including charts and plots, calculating sample statistics, calculating confidence intervals for population parameters, and performing hypothesis testing.

COURSE GOALS:

1. Fundamental Concepts in Probability and Statistical Inference.
2. Counting Methods.
3. Discrete and Continuous Random Variables.
4. Introduction to Statistical Package R.
5. Statistical Inference Topics: Sampling Distributions. Point Estimation. Confidence Intervals. Goodness-of-Fit Testing.

DSA 5103 – Intelligent Data Analysis

INSTRUCTOR: Charles Nicholson

PREREQUISITE: ISE 3293 or ISE 5013; CS 1313 or CS 1323

COURSE DESCRIPTION: Intelligent Data Analytics is an approach to addressing real-world data intensive problems that integrates human intuition with data analysis tools to best draw out meaningful insights. Topics include problem approach and framing, data cleansing, exploratory analysis and visualization, dimension reduction, linear and logistic regression, decision trees, and clustering. Students will be introduced to a powerful open source statistical programming language (R) and work on hands-on, applied data analysis projects.

TEXT: Applied Predictive Modeling by Max Kuhn and Kjell Johnson, 2013. An Introduction to Statistical Learning with Applications in R, 4th printing, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 2013. Recommended Textbook: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition, 10th printing, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman 2009

COURSE TOPICS:

1. Introduction to Analytics.
2. Introduction to R and R Studio.
3. Project Understanding: Describe the phases of the CRISP-DM model. Outline the relationship and interconnectedness of the phases. For the Project Understanding phase, you should be able to: Identify the different steps. For each step, identify the particular requirements and outcomes.
4. Data Understanding: Identify the dimensions of Data Quality. Different Pearson, Spearman and Kendall correlation measures. Explain how Principal Component Analysis works and outline the motivation of using dimension reduction techniques. Run a Principal Component Analysis in R and discuss the results.
5. Data Preparation: Identify a variety of common real-world data quality issues and basic approaches to improve data quality. Determine appropriate approach for outlier treatment. Assess multiple strategies for resolving missing: deletion, single imputation, multiple imputations, maximum likelihood, use R to resolve missing values.
6. Principles of Modeling: Discriminate between Supervised and Unsupervised Learning
7. Multiple Linear Regression.

COURSE GOALS:

1. Defining and framing Analytics Problems.
2. Understanding and coping with Data.
3. Selecting and using appropriate Analytical Tools.
4. Discovering and communicating the Insight.

DSA 5113 – Advanced Analytics and Metaheuristics

INSTRUCTOR: Charles Nicholson

PREREQUISITE: ISE 5103

COURSE DESCRIPTION: Focus is on developing and employing prescriptive analytics – problem solving strategies using advanced methods in the context of Data Science and Analytics for complex, decision problems. Topics include both continuous and combinatorial optimization with an emphasis on problem formulations, traditional techniques (e.g., mathematical programming), and modern heuristics such as simulated annealing, evolutionary algorithms, and swarm optimization. Students will use programming skills to implement algorithms and solve problems.

TEXT: How to Solve It: Modern Heuristics by Zbigniew Michalewicz and David B. Fogel. 2004.

COURSE TOPICS:

1. Introduction to Course; Problem Solving.
2. Introduction to AMPL.
3. Network Flows.
4. Integer Programming Formulation.
5. Branch-and-bound; algorithm complexity.
6. Introduction to heuristics and local search: search space, neighborhoods, encoding.
7. Greedy, Tabu Search, Simulated Annealing.
8. Scatter Search, Grasp, and VNS.
9. Introduction to Python. Introduction to Genetic Algorithms.
10. Genetic Algorithms.
11. Particle Swarm Optimization.
12. Ant Colony Optimization.
13. Multicriteria Decision-Making.
14. Hybrid Methods.

COURSE GOALS:

1. Identification of pros/cons/issues in a variety of problem types and solution approaches.
2. Collaborative development of solution strategies.
3. Implementation of advanced analytics method to solve complex problems.

DSA 5900 – Engineering Professional Practice

INSTRUCTOR: Faculty member

COURSE DESCRIPTION: This course focuses on real-world application of the skills taught in major and concentration courses. This course allows a student to earn credit toward the degree requirements by completing an intense practicum experience.

TEXT: None required

ELIGIBILITY:

The student must first identify a practicum sponsor and secure a practicum topic. The student should present this to the faculty coordinator in his/her school to determine if this practicum meets the requirements for credit. To meet the requirements, the set of activities that the student expects to complete during the practicum must contain higher-level engineering or science work such that the sum of the experiences for the practicum is equivalent to what a student would learn in an equivalently credited, graduate-level course in student's major. In addition, the practicum sponsor overseeing the practicum must agree in writing to prepare two evaluation/progress reports over the course of the experience. The reports can be in the form of an email or letter.

The student must complete a proposal prior to starting the practicum, but after consultation with the faculty coordinator and the practicum sponsor. The proposal should identify the practicum sponsor, the duration of practicum, and the deliverables as agreed upon by the practicum sponsor and the faculty coordinator.

The proposal must include the following sections: Introduction, Objectives, Plan, Deliverables, and Schedule.

COURSE DESCRIPTION:

As a part of the course requirements, the student must complete:

- i) **Written Report:** this report will typically be 15-25 pages long (not including the title page, executive summary, references or appendices), and must be typewritten. The typewritten report should contain all the sections required for the proposal (i.e., Introduction, Objectives, Practicum Plan, and Schedule), updated and expanded as appropriate. The report should also include sections titled "Results" and "Self-Assessment". The Results section should describe the student's responsibilities during the practicum and the scope of work completed. The Self-Assessment section should include the student's assessment of his/her practicum accomplishments following any guidelines provided by the faculty coordinator. The written report must be submitted and presented to the faculty coordinator for review, no later than the last day of classes. It is highly recommended that students submit a draft report to their advisors 2-4 weeks before this date to allow for advisor review and student revisions.
- ii) **Oral Presentation:** a computer-based 20-25 minute oral presentation will be given by the student highlighting the topics as described in the written report. This presentation should be given no later than the last day of classes.