Data Science Program at Montana Tech

The Montana ActiveStatistics Conference
Carroll College · Helena, Montana

Phillip J. Curtiss, Assistant Professor
pcurtiss@mtech.edu, (406) 496-4807

June 28, 2018

Department of Computer Science
Montana Technological University

cs.mtech.edu
www.mtech.edu

Data Science Program Overview

Data Science Program - Collaborative Effort and Goal

The Data Science Program at Montana Tech is a collaborative effort between the Department of Computer Science and the Statistics Department.

The Data Science curriculum is designed to produce graduates with strong statistics and computer science training since a data scientist must be proficient in both areas.

Data Science Program - History

- Originally Data Science Subject Matter was a Concentration Option in the Statistics Program
- Approved by the Board of Regents in September of 2016
- First Course offered during the Fall of 2017
- Currently approximately five (5) undergraduate students in the program and growing
- Statistics Graduates working in Data Science:
  - Chris Dienes, BS Statistics, Montana Tech and Ph.D. Statistics, UC Davis
  - Baili Connors, BS Statistics, Montana Tech

Data Science Program - Additional Courses

- Statistics Department
  - STAT 453 Statistical Learning and Data Science I
  - STAT 454 Statistical Learning and Data Science II
  - STAT 456 Bayesian Statistical Inference
- Department of Computer Science
  - CSCI 347 Data Mining (Refactored)
  - CSCI 444 Data Visualization
  - CSCI 447 Machine Learning
- Capstone Project Senior Year

Statistics Curriculum in Data Science Degree Program

Statistics Curriculum in the Data Science Degree

The goal of the statistics component of the Data Science degree is to give the students a strong background in both theoretical and applied statistics. A data scientist without a strong understanding of statistics if not a well-trained analyst.

STAT 332 Statistics for Scientists and Engineers

Credits: 3 (spring sophomore)

Prereq: M172 Calculus II

Description: Studies probability, random variables, univariate discrete and continuous probability models, expected values, sampling distribution, data collection, point and interval estimation, and hypothesis testing.

STAT 421 Probability Theory

Credits: 3 (fall junior)

Prereq: STAT 332

Description: Studies probability, random variables, expected values, discrete and continuous probability distribution, multivariate probability distributions, generating functions, sums of independent random variables, and the Central Limit Theorem.

STAT 422 Mathematical Statistics

Credits: 3 (spring junior)

Prereq: STAT 421

Description: Studies the theory behind point estimation, tests of hypotheses, confidence intervals, and decision theory. Topics discussed include statistics, order statistics, method of moment estimation, maximum likelihood estimation, decision theoretic estimation, Neyman-Pearson testing, likelihood ratio testing, and confidence interval procedures.

STAT 432 Regression and Model Building

Credits: 3 (spring junior)

Prereq: STAT 332

Description: Detailed study of classical simple linear and multiple regression, model diagnostics and assessment, and an introduction to generalized linear models.

STAT 435 Statistical Computing and Exploratory Data Analysis

Credits: 3 (spring senior)

Prereq: STAT 332

Description: Techniques available to the statistician for efficient use of the computer to perform data management and exploratory statistical analyses. Programming statistical software for the import and export of data, variable sorting, creation of new variables, descriptive statistics and displaying data efficiently are discussed. Special modern statistical topics including bootstrapping and nonparametric modeling are discussed.

STAT 441 Experimental Design

Credits: 3 (fall senior)

Prereq: STAT 332

Description: Studies concepts and applications of the collection of data and proper design of experiments. Aspects of design, sampling principles, exploratory data analysis, confidence intervals and hypothesis testing, and basic experimental design are taught. The importance of randomization, replication, and methods for reducing experimental error are emphasized.

STAT 453 Statistical Learning and Data Science I (new)

Credits: 3 (fall senior)

Prereq: STAT 332 and STAT 432

Description: Introduces the student to statistical learning, model accuracy, simple and multiple linear regression, classification, logistic regression, discriminant analysis, Bayes Theorem for classification, and resampling methods, such as cross-validation and the bootstrap.

STAT 454 Statistical Learning and Data Science II (new)

Credits: 3 (spring senior)

Prereq: STAT 453

Description: Designed to be a rigorous introduction to concepts in statistical learning, focusing on both supervised and unsupervised learning. Topics include linear modeling, non-linear modeling, principle component analysis, clustering methods, and using R.

STAT 456 Bayesian Statistical Inference (new)

Credits: 3 (spring junior)

Prereq: STAT 332

Description: Introduces the student to subjective probability, Law of Total Probability, Bayes Theorem, prior and posterior distribution, conjugate priors, non-informative priors, Bayesian estimation, credible intervals, Bayesian hypothesis testing, prediction, Gibbs sampling, hierarchical Bayesian analysis, and computer implemented data analysis.

Computer Science Curriculum in Data Science Degree Program

Computer Science Curriculum in the Data Science Degree

The goal of the computer science component of the Data Science degree is to give the students a strong background in foundational computer science - including data structures, algorithms, object oriented programming and design and discrete structures. As the students encounter advanced statistical concepts and methods for data science, these concepts and methods are synthesized in the computer science courses of data mining, machine learning, and data visualization.

CSCI 102 Computational Thinking

Credits: 3 (fall freshman)

Prereq: none

Description: Computational thinking involves solving problems, designing systems, and to understanding human behavior, by drawing on the concepts fundamental to computer science. Critical thinking involves the systematic evaluation of information, and is a crucial piece of problem solving. The two are combined to provide a powerful set of tools to understand and solve the kinds of problems they will encounter. The lab incorporates allows students to carefully and systematically analyze problems and demonstrate the correctness of their solution by implementing it in program code.

CSCI 135 Foundations of Computer Science I

Credits: 3 (fall freshman)

Prereq: M151 Precalculus

Description: Introduces the beginning programmer to the fundamentals of object-oriented programming while becoming acquainted with many of the core features of Java. The student will be exposed to the creation of a variety of games and applications. After the basics, the student will learn to develop his or her own systems of classes, eventually working with many of Java's Graphical User Interface (GUI) features and will develop a desktop Windows application.

CSCI 136 Foundations of Computer Science II

Credits: 3 (spring freshman)

Prereq: CSCI 135

Description: A continuation of CSCI 135 . Explores the finer and more subtle elements of a programming language. Programming techniques and structures include arrays, records, and linked lists.

CSCI 232 Data Structures and Algorithms

Credits: 3 (fall sophomore)

Prereq: CSCI 136

Description: Commonly used structures found in computing and the algorithms which manipulate them are studied. Design and analysis of algorithms are emphasized. Topics include stacks, queues, general lists, trees, hashing, searching, and sorting.

CSCI 246 Discrete Structures

Credits: 3 (fall sophomore)

Prereq: CSCI 136

Description: Course includes those mathematical topics which will help students in future courses. It refines problem solving skills by providing a vocabulary, structures and techniques for working with problems. Topics include logic, theorem proving, properties of sets, functions, relations, and sequences, counting techniques, recursion, and graph theory.

CSCI 332 Design and Analysis of Algorithms

Credits: 3 (spring sophomore)

Prereq: CSCI 232

Description: Uses and reinforces basic data structure knowledge and techniques from Data Structures and Algorithms (CSCI 232). Covers several advanced data structures, including balanced search trees and graphs. Studies common algorithm design methods (Brute Force, Decrease and Conquer, Divide and Conquer, Greedy, and Dynamic Programming) to solve various classic problems. Emphasizes the space and time complexities of various data structures and their associated algorithms.

CSCI 340 Database Design

Credits: 3 (spring sophomore)

Prereq: CSCI 136

Description: Studies concepts and applications in database management including the relational model, relational algebra, Structural Query Language (SQL), normalization, transactions, and how to avoid SQL-injection. XML, No SQL and Big Data are introduced. Information stewardship is discussed. Students get a chance to prototype a database application working in groups.

CSCI 347 Data Mining (refactored)

Credits: 3 (fall junior)

Prereq: CSCI 340

Description: Provides a grounding in data mining techniques and prepares students to design, use, and evaluate these techniques on a variety of application domains and for the purpose of decision support. Topics include decision trees, rule based systems, statistical approaches, neural networks, and instance-based approaches.

CSCI 444 Data Visualization (new)

Credits: 3 (spring junior)

Prereq: CSCI 135

Description: Covers the principles, methods, and techniques for effective visual analysis of data. Students will use both common and special software packages to explore different visualization applications. Students will learn how to formulate 3-D numerical models, translate 3-D models into graphical displays, and create time sequences and pseudo-animations. The course covers interactive versus presentation techniques and special techniques for video, DVD, and other media displays.

CSCI 446 Artificial Intelligence

Credits: 3 (fall senior)

Prereq: CSCI 332

Description: An introduction to the basic concepts of Artificial Intelligence. Topics to be covered include the history of AI, the problems treated in AI, solution techniques, state spaces, search algorithms and heuristics, expert systems, natural language processing, and robotics.

CSCI 447 Machine Learning (new)

Credits: 3 (spring senior)

Prereq: CSCI 332

Description: Introduction to the framework of learning from examples, various learning algorithms such as neural networks, and generic learning principles such as inductive bias, Occam's Razor, and data mining. Reviews some statistical learning techniques, but focus is on non-statistical techniques.

Capstone Project Requirement for Data Science Degree Program

STAT/CSCI 499 Capstone Project

Credits: 4 (spring senior)

Prereq: Data Science Major and Senior Standing

Description: Data science concepts and methods applied to a specific problem domain.

Data Science Student Opportunities

Data Science Student Opportunities - Research

- Summer Undergraduate Research Fellowship (SURF)
- Undergraduate Research Program (AY)

Data Science Student Opportunities - Internships

- Micron in Boise Idaho - 2 Students
- White Sands New Mexico - 1 Student
- Google - 2 Students

Data Science Student Opportunities - Employment

- Micron in Boise Idaho - Manufacturing, Process
- Sofi in Helena - Analytics, Financial Modeling

Challenges thus Far

Challenges

- Recruiting into the Program
- Faculty Resources
- Data Analytics

Challenges - Strategies for Addressing

- Recruiting into the Program - outreach coordinator and university engagement
- Faculty Resources - data science experiences, background in new hires
- Data Analytics - possible collaboration with business information technology

Questions?

For Additional Information

- Rick Rossi, Ph.D.
  Program Director
  Statistics Department
  406-496-4356, rrossi@mtech.edu

- Michele Van Dyne, Ph.D.
  Department Chair
  Department of Computer Science
  406-496-4855, mvandyne@mtech.edu

Montana Tech Web site at http://www.mtech.edu
Dept. of Computer Science Web site at http://cs.mtech.edu