

Proposed New Minor in Data Science

15 credits

1. DS 151 Introduction to Data Science
2. DS 261 Ethics and Privacy in Data Science
3. MATH 474 Probability and Statistics or MATH 476 Statistics
4. MATH 484 Regression or CS 422 Data Mining or CS 484 Introduction to Machine Learning
5. DS 451 Data Science Life Cycle or CSP 571 Data Preparation and Analysis

Course Descriptions

DS 151 Introduction to Data Sciences

This course intends to introduce the critical concepts and skills in statistical inference, machine learning, and computer programming, through hands-on analysis of real-world datasets from various fields. Simplified, open-source data analysis and visualization tools are available that form a stepping stone to industry-standard programming languages, which broadens the accessibility of this course beyond programming aficionados.

Lecture: 3 Lab: 0 Credits: 3

DS 261 Ethics and Privacy in Data Science

This course takes a technical approach to explore societal issues of ethics, fairness, responsibility, and privacy related to the collection, use, and generalization of data. Technical measures of fairness and equity are presented, along with their potential benefits and shortcomings. Statistical and algorithmic techniques for providing data privacy are discussed and critiqued in real-world contexts. Case studies are used to expose breakdowns in the data science pipeline that have led to inequities or actual harm, requiring mitigation or refactoring of the pipeline. Specific focuses include the effect of uneven data collection or misuse of data generalization on underrepresented groups and built-in algorithmic or statistical bias.

Prerequisite(s): DS 2XX Mathematical Foundations for Data Science.

Lecture: 3 Lab: 0 Credits: 3

DS 451 Data Science Life Cycle

Stages of a data science project from start to finish: obtaining data, exploring data, determining what questions the data can answer, exploratory analysis, ethical impacts analysis and mitigation, hypothesis (re-)formulation, in-depth analysis, validation, and reporting. Case studies extensively included. Designed to educate the DS student in the typical project life-cycle stages required in the data science professions.

Prerequisite(s): DS 261 Ethics and Privacy in Data Science and one of DS 484 Introduction to Statistical Learning/MATH 484 Regression or CS 484 or CS 422.

Lecture: 3 Lab: 0 Credits: 3

Math 474 Probability and Statistics

Elementary probability theory including discrete and continuous distributions, sampling, estimation, confidence intervals, hypothesis testing, and linear regression. Credit not granted for both Math 474 and Math 475.

Prerequisite(s): Math 251 or Graduate standing.

Lecture: 3 Lab: 0 Credits: 3

Math 476 Statistics

Estimation theory; hypothesis tests; confidence intervals; goodness-of-fit tests; correlation and linear regression; analysis of variance; nonparametric methods.

Prerequisite(s): Math 475 or Graduate standing.

Lecture: 3 Lab: 0 Credits: 3

Math 484 Regression

This course introduces the basic statistical regression model and design of experiments concepts. Topics include simple linear regression, multiple linear regression, least square estimates of parameters; hypothesis testing and confidence intervals in linear regression, testing of models, data analysis and appropriateness of models, generalized linear models, design and analysis of single-factor experiments.

Prerequisite(s): Math 474 with min. Grade of C or (Math 476 with min grade of C and Math 475 with min grade of C)

Lecture: 3 Lab: 0 Credits: 3

CS 422 Data Mining

This course will provide an introductory look at concepts and techniques in the field of data mining. After covering the introduction and terminologies to Data Mining, the techniques used to explore the large quantities of data for the discovery of meaningful rules and knowledge such as market basket analysis, nearest neighbor, decision trees, and clustering are covered. The students learn the material by implementing different techniques throughout the semester.

Prerequisite(s): CS 331 or CS 401 or CSSP 401 with min. .grade of S or CS 403

Lecture: 3 Lab: 0 Credits: 3

CS 484 Introduction to Machine Learning

An introduction to machine learning concepts and algorithms, including classification, clustering, and regression. Topics include k-means clustering, nearest neighbors classification, decision trees, naive Bayes, logistic regression, support vector machines, and neural networks. Special focus will be on practical aspects of machine learning, including data preparation, experimental design, and modern tools for building machine learning systems. Basic probability theory knowledge is required.

Prerequisites(s): Math 151 and CS 116 or CS 201 or CS 401

Lecture: 3 Lab: 0 Credits: 3

CSP 571 Data Preparation and Analysis

Surveys industrial and scientific applications of data analytics with case studies including exploration of ethical issues via case studies. Students will work with a variety of real world data sets and learn how to prepare data sets for analysis by cleaning and reformatting. We will also cover a variety of data exploration techniques including summary statistics and visualization methods.

Prerequisite(s): (CS 331 or CS 401 or CSSP 401) and CS 425 and Math 474

Lecture: 3 Lab: 0 Credits: 3