DATA SCIENCES (DS)

DS 97: SPECIAL TOPICS

1-9 Credits/Maximum of 9

Formal courses given infrequently to explore, in depth, a comparatively narrow subject that may be topical or of special interest.

DS 99: Foreign Studies

1-12 Credits/Maximum of 999

Courses offered in foreign countries by individual or group instruction.

International Cultures (IL)

DS 120: Scripting for Data Sciences

1 Credits

Introductory course in computer-based scripting languages for use in data analyses. DS 120 Scripting for Data Sciences (1) This introductory course aims to teach practical skills in data manipulation and preprocessing scripting, including the fundamentals of an interpreted programming language for use in the data sciences. The goal of the course is to provide an accessible (no pre-requisites) and brief (1 credit) introduction, preparing students for hands-on data analytics assignments in DS 200 Introduction to Data Sciences. This practical course teaches fast manipulation of datasets on the Unix command line, scripting in spreadsheets, and fundamental control structures and data manipulation in a modern interpreted programming language. It is expected that students gain an overview of the available tools and techniques that allows them to acquire basic proficiency in select techniques in the course of applications in most other courses in Data Sciences.

DS 197: Special Topics

1-9 Credits/Maximum of 9

Formal courses given infrequently to explore, in depth, a comparatively narrow subject that may be topical or of special interest.

DS 199: Foreign Studies

1-12 Credits/Maximum of 12

Courses offered in foreign countries by individual or group instruction.

DS 200: Introduction to Data Sciences

3 Credits

Introduction to Data Sciences is an introductory survey course appropriate to a wide range of students. The course provides an overview of topics covered in more detail in advanced Data Sciences Courses. The course is multidisciplinary in nature, covering technical subjects such as basic aspects of machine learning, analyzing Big Data and visual analytics, but also examining the managerial aspects of data including data management and decision making, in addition to organizational and societal implications of increasing data collection and processing and their implications for privacy and security.

DS 220: Data Management for Data Sciences

3 Credits

This course has two components: (1) advancing students¿ knowledge on relational database and (2) introducing NoSQL databases. In the first component, the course will review the techniques learned in IST 210, strengthen students¿ skills in using SQL queries and introduce students about indexing and scalability issue in relational database. While relational database is still frequently used, the emergence of storage for big data and various types of data has driven a new of class of non-relational databases commonly referred to NoSQL database. This course will introduce the real-world needs for NoSQL databases and the characteristics that distinguish them from relational database. We will introduce both the concepts of NoSQL databases and how the concepts are implemented in the database systems. We will focus on three main NoSQL data models: key-value, column family, and document. Students will learn the concepts of these data models and know how to use them in the database systems. We will also introduce the concepts on graph database, hadoop system, and warehousing. Finally, we will present criteria that decision makers should consider when choosing between relationship between relational and non-relational databases and techniques for selecting the NoSQL database that best addresses specific use cases.

Enforced Prerequisite at Enrollment: CMPSC 121 and IST 210

DS 294: Research Project

1-12 Credits/Maximum of 12

Supervised student activities on research projects identified on an individual or small-group basis.

DS 296: Independent Studies

1-18 Credits/Maximum of 18

Creative projects, including research and design, that are supervised on an individual basis and that fall outside the scope of formal courses.

DS 297: Special Topics

1-9 Credits/Maximum of 9

Formal courses given infrequently to explore, in depth, a comparatively narrow subject that may be topical or of special interest.

DS 299: Foreign Studies

1-12 Credits/Maximum of 12

Courses offered in foreign countries by individual or group instruction.

International Cultures (IL)

DS 300: Privacy and Security for Data Sciences

3 Credits

The course provides students with the knowledge and skills to analyze and implement protection strategies for data privacy and security.

Enforced Prerequisite at Enrollment: DS 220

DS 310: Machine Learning for Data Analytics

3 Credits

The course teaches students the principles of machine learning (and data mining) and their applications in the data sciences. DS 310 Machine Learning for Data Analytics (3) The course introduces the principles of machine learning (and data mining), representative machine learning algorithms and their applications to real-world problems. Topics to be covered include: principled approaches to clustering, classification, and function approximation from data, feature selection and dimensionality reduction, assessing the performance of alternative models, and relative strengths and weaknesses of alternative approaches. The course will include a laboratory component to provide students with hands-on experience with applications of the algorithms to problems from several domains. Prerequisites for the course include basic proficiency in programming, elementary probability theory and statistics, and discrete mathematics.

Enforced Prerequisite at Enrollment: CMPSC 121 and STAT 318

DS 320: Data Integration

3 Credits

Recommended Preparations: DS 310 Modern data-intensive applications (healthcare, security, public policy, science, commerce, crisis management, education, among others) increasingly call for integration of multiple types of data from disparate sources. This course introduces students to the principles and the practice of data integration, with particular emphasis on relational, knowledge-based, graph-based, and probabilistic methods. Carefully crafted assignments will help enhance the students¿ mastery of both the theoretical underpinnings as well as practical aspects of data integration. The students will work in teams to solve representative data integration problems drawn from real-world applications. Upon completion of the course, students should be able design, implement, and evaluate data integration solutions to support data intensive applications.

Enforced Prerequisite at Enrollment: DS 220 and STAT 318.
Recommended Preparation: DS 310

DS 330: Visual Analytics for Data Sciences

3 Credits

The course introduces visual analytics methods and techniques that are designed to support human analytical reasoning with data. DS 330 Visual Analytics for Data Sciences (3) Visual analytics is the science of combining interactive visual interfaces and information visualization techniques with automatic algorithms to support analytical reasoning through human-computer interaction. People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data, and to communicate their findings effectively for decision-making. This course will serve as an introduction to the science and technology of visual analytics and will include lectures on both theoretical foundations and application methodologies. The goals of this course are for students to (1) develop a comprehensive understanding of this emerging, multidisciplinary field, and (2) apply that understanding toward a focused research problem in a real-world application or a domain of personal interest.

Enforced Prerequisite at Enrollment: DS 220

DS 340W: Applied Data Sciences

3 Credits/Maximum of 3

This course builds up the students¿ understanding of data sciences by discussing the fundamental principles in the context of real-world examples, and then shows specifically how the principles can provide understanding of many of the most common methods and techniques covered in previous data science courses. The course features three individual projects as well as a team project spanning the entire course. After taking this course, the students should be able to cover the entire pipeline of a data science project, from problem formulation to data science solutions. That is, start from a data driven problem, identify pertinent datasets to the problem and collect data, reason about the best techniques that should be used to solve the problem, implement algorithms and models, assess performance, and communicate actionable insights through both written reports and oral presentations. As one example, a fundamental principle of data science is that solutions for extracting useful knowledge from data must carefully consider the problem in the real world scenarios. This may sound obvious at first, but the notion underlies many choices that must be made in the process of data analytics, including problem formulation, method choice, solution evaluation, and general strategy formulation. Another fundamental principle is that predictive modeling can both inform and be informed by relevant knowledge (including theories, models, frameworks) of the relevant domains. This principle manifests itself throughout data science: in the specific design of many particular data sciences applications, and more generally as the basis for all ¿intelligent¿ solutions. In this course, this principle will be highlighted by case studies from multiple domains so that students can be inspired to apply this principle to their term projects. Lastly, as most data science projects are delivered as solutions as opposed to software deliverables, the ability for data scientists to communicate their results through concise and actionable insights plays a critical role in a data science project. This course places a particular focus on developing student writing abilities, through formal project reports and presentations. The individual projects will offer an interactive experience for students through feedbacks on their reports provided by the instructor. The term-long project will also train students in writing in a collaborative environment.

Enforced Prerequisite at Enrollment: DS 300 and DS 310 or CMPSC 448.
Recommended Preparation: DS 330.
Writing Across the Curriculum

DS 396: Independent Studies

1-18 Credits/Maximum of 18

Creative projects, including research and design, that are supervised on an individual basis and that fall outside the scope of formal courses.

DS 397: Special Topics

1-9 Credits/Maximum of 9

Formal courses given infrequently to explore, in depth, a comparatively narrow subject that may be topical or of special interest.

DS 399: Foreign Studies

1-12 Credits/Maximum of 12

Courses offered in foreign countries by individual or group instruction.

International Cultures (IL)

DS 402: Emerging Trends in the Data Sciences

3 Credits/Maximum of 9

This course exposes and trains students in the analysis of emerging trends in data sciences. DS 402 Emerging Trends in the Data Sciences (3) Data sciences is a rapidly evolving field affected by innovations in a variety of technical domains, including data generation, capture, storage, and processing. Staying abreast of new developments can be a daunting task but is critical for success. This course provides an in-depth analysis of a particular innovation, but starts with developing generally applicable skills for analyzing new technologies. In particular, the analytic framework considers the innovation’s technical aspects and potential for widespread adoption, but also its social, organizational and policy implications. As a course focused on a new data sciences technology or analytic innovation, it is repeatable. As such, the course enables students to be exposed to the cutting edge of data sciences, supporting a forward looking view of the field for students across the university.

Enforced Prerequisite at Enrollment: DS 220

DS 410: Programming Models for Big Data

3 Credits

Recommended Preparations: DS 310; CMPSC 448 This course introduces modern programming models and related software stacks for performing scalable data analytics and discovery tasks over massive and/or high dimensional datasets. The learning objectives of the course are that the students are able to choose appropriate programming models for a big data application, understand the tradeoff of such choice, and be able to leverage state-of-the art cyber infrastructures to develop scalable data analytics or discovery tasks. Building on data models covered in DS 220, this course will introduce programming models such as MapReduce, data flow supports for modern cluster computing environment, and programming models for large-scale clustering (either a large number of data samples or a large number of dimensions). Using these frameworks and languages, the students will learn to implement data aggregation algorithms, iterative algorithms, and algorithms for generating statistical information from massive and/or high-dimensional data. The realization of these algorithms will enable the students to develop data analytic models for massive datasets.

Enforced Prerequisite at Enrollment: CMPSC 122 and DS 220.
Recommended Preparation: DS 310 or CMPSC 448.
Cross-listed with: CMPSC 410

DS 440: Data Sciences Capstone Course

3 Credits/Maximum of 3

This course provides a data sciences problem-solving experience, addressing realistic data science dilemmas for which solutions require teamwork and collaboration.

Enforced Prerequisite at Enrollment: DS 220. Recommended preparation: DS 310 or CMPSC 448.

DS 494: Research Project

1-12 Credits/Maximum of 12

Supervised student activities on research projects identified on an individual or small-group basis.

DS 496: Independent Studies

1-18 Credits/Maximum of 18

Creative projects, including research and design, that are supervised on an individual basis and that fall outside the scope of formal courses.

DS 497: Special Topics

1-9 Credits/Maximum of 9

Formal courses given infrequently to explore, in depth, a comparatively narrow subject that may be topical or of special interest.