

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 197G: Special Topics - GenEd

1-9 Credits/Maximum of 9

Formal courses given on a topical or special interest subject offered infrequently; several different topics may be taught in one year or semester. This Special Topics is a GenEd course.

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

4 Credits

The course introduces students to data sciences, an emerging discipline focused on the knowledge and skills needed to harness the power of data to advance science and engineering, address complex national and global challenges, inform public policy, and improve human lives. It demonstrates how the discipline of data science integrates knowledge and skills in computer sciences, statistics, and informatics (with

exposure to application domains such as life science, health science, cyber security, astronomy, etc). Through a combination of lectures, hands-on labs, and case studies, students are introduced to the "big picture" of data sciences including elements of understanding data through exploratory data analysis, testing hypotheses against data, building predictive models, all using real-world examples. The course also introduces students to opportunities to specialize in Applied Data Sciences (with an emphasis on data sciences applications in the real world), Computational Data Sciences (with an emphasis on well-engineered data analytics systems), and Statistical Data Sciences (with an emphasis on advanced statistical theory and methods).

DS 220: Data Management for Data Sciences

3 Credits

The course introduces students to the fundamentals of data models: organizing, managing, and using different types of data that arise in real-world applications. The course introduces students to several alternative data models and database solutions, emphasizing their strengths and limitations in the context of real-world applications. Topics covered include the relational databases, key-value stores, column-oriented databases, vector-space databases, graph databases, and distributed file systems together with their applications in solving real-world big data management problems. Upon completion of the course, the students will be able to choose an appropriate data model and database solution for a given application, and use the chosen database to organize, manage, and use data in the context of specific applications.

Enforced Prerequisite at Enrollment: C or better in CMPSC 121 or CMPSC 131
Enforced Concurrent at Enrollment: CMPSC 122 or CMPSC 132

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

This course provides an overview about data privacy and security implications arising in the context of data analytics, as well as the techniques and processes for managing privacy and security of large-scale structured and unstructured data sources. Concepts are taught in reference to a variety of data types and application areas. Students will learn design principles to enhance privacy and security in data science practice.

Enforced Prerequisite at Enrollment: C or better in (DS 200 or STAT 200) and DS 220 and (IST 230 or CMPSC 360 or MATH 311W)

DS 305: Algorithmic Methods and Tools

3 Credits

This course teaches students how to formulate data science problems that arise in different applications that involve different types of data (tabular, sequence, network, matrix data); and introduces students to common strategies for formulating, and solving those problems. The course will cover (1) how to formulate data science problems (e.g., text analysis, biological sequence analysis, social network analysis, recommender systems) and how to evaluate alternative formulations, (2) common algorithmic methods and tools (as implemented in software libraries) for representing, processing, and sampling data, (3) how to apply the methods and tools to solve well-formulated data science problems. The course will also teach students how to understand and use results about the correctness, efficiency, and scalability of the techniques and tools, and they will learn to recognize when specific algorithmic tools can be used to improve the performance of their data science tasks. Through exercises, students will gain hands-on experience in problem formulation and solution development for data science problems that arise in applications.

Enforced Prerequisite at Enrollment: Prerequisites: (CMPSC 122 or CMPSC 132 or IST 242) and (IST 230 or CMPSC 360 or MATH 311W)

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: C or better in (CMPSC 122 or CMPSC 132) and (STAT/MATH 318 or STAT/MATH 414 or STAT/MATH 418) and MATH 220 and (IST 230 or CMPSC 360 or MATH 311W)
Recommended Preparation: DS 305

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 Prerequisites at Enrollment: DS 220 and (STAT/MATH 318 or STAT/MATH 414 or STAT/MATH 418)

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: C or better in (DS 200 or STAT 200) and (CMPSC 122 or CMPSC 132 or IST 242)

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 Prerequisites at Enrollment: (CMPSC 122 or CMPSC 132) and (DS 220 or CMPSC 221)

Cross-listed with: CMPSC 410

DS 420: Network Analytics

3 Credits

Networks constitute a useful representational abstraction for data from many real-world applications, such as social networks, biomolecular networks, brain networks, among others. This course aims to cover the conceptual, algorithmic, and applied aspects of network analytics to prepare the students to analyze network data. Specific topics to be covered include different kinds, e.g., undirected, directed, weighted, unweighted, labeled, unlabeled, networks; network properties and network statistics; statistical models of static and dynamic networks; classes of networks, different classes of network structure, e.g., random, small-world, modular hierarchical; methods for community detection, information propagation, node and link classification, network embedding, and their applications in life sciences e.g., predicting protein function, social sciences, e.g., analyzing social ties, e-commerce e.g., recommender systems.

Enforced Prerequisites at Enrollment: DS 220 and DS 310

DS 435: Ethical Issues in Data Science Practice

3 Credits

This course explores social and ethical dimensions of data science. Datafication can be a powerful force for good, but it can also do harm to individuals and society. Oriented primarily around case studies, the course investigates when, why, and how data is collected, analyzed, and used, and explores the ethical stakes of data-driven systems. In addition to diagnosing ethical problems-e.g., invasions of privacy, algorithmic bias, and lack of transparency and accountability-students are asked to think creatively and constructively about how the tools of data science can be used to realize shared ethical and social commitments. The course will be comprised of both "theory" and "lab" components. The former will contextualize ethical problems, introducing students to ethical theories and frameworks for addressing them. The latter will ask students

to put those ideas to work, using the tools of data science to identify examples of ethical issues in data science practice, and proposing means of addressing them.

Enforced Prerequisite at Enrollment: DS 220

DS 440: Data Sciences Capstone Course

3 Credits/Maximum of 3

This course provides opportunities for students to develop their data science problem solving skills via a semester long group project. Projects typically involve analyses of real data. Students should be able to conceptualize a research project, plan its execution, implement the plan, and communicate their results. The capstone projects will integrate knowledge gained in technical subjects such as machine learning, data mining, data integration, and visualization. Other aspects of problem solving including considerations of security, privacy, fairness and ethical issues will often be required. At the end of the semester, students will communicate the results of their projects through a written report and an oral project presentation.

Enforced Prerequisite at Enrollment: C or better in DS 220 and DS 340W
Enforced Concurrent at Enrollment: DS 435 and (DS 305 or CMPSC 465)

DS 440W: Data Science Capstone

3 Credits

The Data Sciences capstone develops the research orientation and creative problem solving necessary for successful careers. The capstone develops these skills in the context of a semester long project, the solution to which requires integration of knowledge, skills and analytic techniques taught in the core curriculum. The Data Sciences capstone will also give student a real world experience in which they will need to work in teams and will be coached on ways to translate analytic outcomes into meaningful and actionable information for decision makers. The course is intended for seniors who have successfully completed the core courses. The capstone projects will integrate knowledge gained in technical subjects such as machine learning, data mining, data integration and visualization. The project will also require students to analyze the privacy and security aspects of the problem as well as apply concepts of decision analysis. Students will also hone their presentation and technical writing skills, generating effective reports that not only explain their analytic processes, assumptions underlying the processes and outcomes, but also communicate the limitations of their approach and potential alternate strategies.

Enforced Prerequisite at Enrollment: Data Science Major or 7th semester standing
Writing Across the Curriculum

DS 441: Information Retrieval and Organization

3 Credits

The practices and foundations of access to textual and nontextual information using the principles of information retrieval and web search. Introductory course for for undergraduate students in the last year of their academic program and graduate students covering the practices, issues, and theoretical foundations of organizing and analyzing information and information content for the purpose of providing access to textual and nontextual information resources. Introduces students to the principles of information storage and retrieval systems and databases. IST 441 Information Retrieval and Organization (3) This is

an introductory course for Information Sciences and Technology senior and graduate students covering the practices, issues, and theoretical foundations of organizing and analyzing information and information content for the purpose of providing access to textual and non-textual information resources. This course will introduce students to the principles of information storage and retrieval systems and databases. Students will learn how effective information search and retrieval is interrelated with the organization and description of information to be retrieved. Students will also learn to use a set of tools, such as search engines, and procedures for organizing information. They will become familiar with the techniques involved in conducting effective searches of information resources.

Enforced Prerequisite at Enrollment: C or better in MATH 141 and DS 220 and (IST 230 or CMPSC 360 or MATH 311W)

DS 442: Artificial Intelligence

3 Credits

This course provides an overview of the foundations, problems, approaches, implementation, and applications of, artificial intelligence. Topics covered include problem solving, goal-based and adversarial search, logical, probabilistic, and decision theoretic knowledge representation and inference, decision making, and learning. Through programming assignments that sample these topics, students acquire an understanding of what it means to build rational agents of different sorts as well as applications of AI techniques in language processing, planning, vision.

Enforced Prerequisite at Enrollment: CMPSC 221. Enforced Concurrent at Enrollment: CMPSC 465
Cross-listed with: CMPSC 442

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.

DS 499: Foreign Studies

1-12 Credits/Maximum of 12

Courses offered in foreign countries by individual or group instruction

International Cultures (IL)