INDIANA UNIVERSITY
DATA SCIENCE

Data Science Graduate Degree Handbook

For the Master of Data Science Degree

2021 – 2022
April 15, 2021

Table of Contents

1. Vision and Mission of Data Science Program 3
What is Data Science? 4
Difference between Data Science and Computer Science? 4
Summary of Luddy's Master of Data Science Degree 4
Data Science is STEM 5
Program outcomes: what are the learning outcomes of the MS Data Science degree? 5
Length of Program 5
Academic Expectations 5
Career Outlook 5
Luddy Career Services 6
Student Responsibilities & Professional Ethics 6
Commitment to Diversity and Inclusion 6
Student organizations 7
Prerequisites and Remedial Coursework 7
Curriculum 8
Internship Course Credits 8
2. Applied Data Science Track Overview and Track Requirements 9
3. Big Data Systems Track Overview and Requirements 16
4. Computational and Analytical Track Overview and Requirements 21
5. Managerial Data Science Track Overview and Requirements 25
6. Other Resources 31
Course Substitution and Exceptions 31
Grades 31
Grade Appeals 31
Academic Probation 31
Registering for Classes 31
How to Register for courses and enrollment at the shopping center? 32
Class Permissions 32
Waitlist 32
Class Withdrawal 33
Bursar Bill 33
Graduation 33
Diplomas 33
Transcripts 33
Funding 33
Full-time Status and Part-time Status 34
Leave of Absence 34
Important Student Resources 34
Luddy Graduate Studies Office Contact 34

INDIANA UNIVERSITY
LUDDY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

2020-2021 Handbook for the Master of Data Science Degree

1. Vision and Mission of Data Science Program

Improve the Way the World Uses Data

As a student of the Data Science Program, you'll prepare to become the data professional our evolving world needs: a holistically trained expert with the vision and skills to use data to solve problems, unite communities, prevent disasters, transform industries, and most importantly, improve lives.

The Master of Data Science program gives our students a deep set of core competencies to see what tomorrow can be, and shape it every day:

- Prepare students for courses in statistics, ML, data management and engineering
- Empower students to apply computation and inferential thinking to tackle real-world problems
- Enable students to start careers as data scientists by providing experience with data tools and techniques

Director, Data Science Academic Programs
Haixu Tang
Professor of Informatics & Computing

Director of Graduate Studies for Data Science
Patrick C. Shih
Assistant Professor of Informatics

What is Data Science?

Data science is the mining, collecting, analyzing, managing, and storing data to help make data driven decisions in e-commerce, finance, government, healthcare, science, social networking, telecommunications, politics, utilities, smart meters, education, aerospace, etc. By collecting, analyzing, managing, and storing data, businesses can run more efficiently and make data-driven business decisions.

To prepare for a career in data science, students need to be proficient in math, statistics, and computer programming such as Python or R. Students need to understand the data in order to analyze and interpret the data in a meaningful way. To visualize the data, data scientists often use Tableau, Hadoop, or Apache Spark.

According to KDNuggets, “data scientists are highly educated – 88% have at least a master’s degree.” Their undergraduate background is in computer science, statistics, social science, or physical science.

Difference between Data Science and Computer Science?

The main difference between a data scientist and a computer scientist is that a computer scientist develops software and data scientists use the software developed by computer scientists to analyze and interpret the data and identify trends.

If you like to mine, collect, analyze, manage, and store data, perhaps you should pursue a master’s degree in data science degree as data scientists mine, collect, analyze, manage, and store data to help make data driven decisions. Data scientists have a good understanding of the data by asking and answering questions as they do their analysis. They are adept in pulling data from multiple sources, cleaning up the data, and analyzing the data to help make sound business decisions. By analyzing the data, the data scientist can make suggestions as to how to improve the process or how to make the process more efficient. When presenting the data to stakeholders, the data scientist designs, creates, and builds data models and data visualizations to make the data easier to understand.

To be competitive in the job market, a large majority of companies are looking for students who have a bachelor’s degree coupled with a master’s degree in data science. The most common data science job titles are data scientist, data architect, data engineer, business analyst, or data analyst.
If you like to build new things, perhaps you should pursue a master’s degree in computer science as computer scientists design, create, test, document, and debug code, software, and mobile applications. Often computer scientists collaborate with other computer scientists and their teams in developing a larger piece of software, application, or computer system.

To be competitive in the computer science job market, you need at least a bachelor’s degree in computer science. Students who have a master’s degree in computer science are paid more, have more responsibility, and more room for advancement in a company. The most common computer science job is software development engineer, software developer, Java developers, systems engineer, or network engineer.

It is expected that students who have degrees in data science and computer science will be in high demand for at least the next five to ten years. By earning a master’s degree in data science or computer science coupled with your undergraduate degree will give you an edge on the job market.

Summary of Luddy’s Master of Data Science Degree

The Master of Data Science degree is interdisciplinary in computer science, information science, informatics, statistics, engineering, and other disciplines. It prepares students to pursue a data science related career as a data scientist, data analyst, data architect, etc. or admission to a Ph.D. program.

To earn the Master of Data Science degree, you must successfully earn 30 graduate-level credit hours. The program takes two years to complete. As a Master of Data Science student, you have the option of focusing on one of the following four distinct tracks: (1) Applied Data Science; (2) Big Data Systems; (3) Computational and Analytical; and (4) Managerial Data Science.

Data Science is STEM

Data Science is in the STEM field (science, technology, engineering, or mathematics). Since the Data Science program is interdisciplinary and an applied program, international students are eligible for a STEM OPT Extension. For more information about the STEM OPT Extension and a list of qualifying STEM majors, go to https://ois.iu.edu/living-working/employment/f1/optional/stem-opt.html.

Program outcomes: what are the learning outcomes of the MS Data Science degree?

The Data Science program gives our students a deep set of core competencies in multiple areas—including programming, statistics, data analytics, machine learning, data wrangling, data visualization, communication, business foundations, and ethics that increase their marketability in the industry. The learning outcomes of the MS Data Science Residential degree are the knowledge and skills acquired in the program that are transferable to successfully use data to solve problems, which include:

- Data preparation and presentation
- Exploratory data analytics & visualization
- Model fitting and inference
- Efficient and scalable data processing

Length of Program

The Master of Data Science degree requires a student to successfully complete 30 credit hours. Master's students must be enrolled full-time each semester. Typically, it takes students two years to complete the Master of Data Science program.

During the first three semesters, students take nine (9) credit hours per semester and three (3) to nine (9) credit hours during the fourth semester. The student's advisor, program director, and the Director of Graduate Studies must approve exceptions. During the summer between Year I and Year II of their studies, students often take an internship.

Academic Expectations

We expect students to develop as a scholar, an instructor-mentor, and a professional. As a master's student and in your career, it is expected that students maintain professionalism and high standards in your interactions with faculty, staff, colleagues, and students as well as in your role as a researcher or associate instructor.

All students must (1) maintain cumulative and semester GPAs of 3.0 or above; (2) complete coursework in a timely manner; (3) maintain academic integrity; (4) maintain a good academic standing; and (5) conduct themselves in accordance with the Indiana University's Code of Student Rights, Responsibilities, & Conduct (http://www.indiana.edu/~code/). Failure to maintain any of the above requirements will result in the student being placed on academic probation or dismissal from the program. Funding may be in jeopardy as well.

Career Outlook

Data Science is shaping the future. According to the U.S. Bureau of Labor Statistics Report, by 2026, there will be 11.5 million job openings in the field of Data Science. According to Dr. Martin Schedlbauer, a

Data Science Professor at Northeastern University, “data science careers are in high demand and this trend will not be slowing down any time soon, if ever.”

The demand for data scientists is high. With a Master of Data Science degree from Indiana University’s Luddy School of Informatics, Computing, and Engineering, you could pursue a career as a:

- Business Intelligence Developer
- Data Architect
- Data Scientist
- Data Analyst
- Data Engineer
- Decision Scientists
- Enterprise Architect
- Software Developers
- Statistician

Luddy Career Services

The Luddy School of Informatics, Computing, and Engineering’s Office of Career Services offers a variety of programs and services to help students find and succeed in internships and full-time jobs. The Office of Career Services will review student’s resumes and cover letters, will hold mock interviews, will assist in negotiating a hiring package, etc.

The Indiana University’s Career Development Center is also available to Luddy graduate students. In the fall and spring, the Luddy Office of Career Services hosts two large career fairs. Many of the employers who attend these career fairs are looking to hire students for full-time employment or internships. For Luddy Career Outcomes, go to our Career Services Website.

Student Responsibilities & Professional Ethics

All students must abide by the Indiana University Code of Student Rights, Responsibilities, & Conduct (http://www.indiana.edu/~code/). This applies to scholarship, any role the student may have as an Associate Instructor (AI), relations with colleagues, relations with students, and compliance with academic standards with respect to academic ethics.

If students are not familiar with the concept and best practices of avoiding any hint of plagiarism in American universities, they should become familiar with these standards. The Code provides a series of documents describing the behaviors, ideals, and goals for Indiana University.

Commitment to Diversity and Inclusion

Our commitment to diversity, equity, and inclusion is grounded in our aspiration to cultivate intellectual rigor and curiosity among our students and to prepare them to thrive in and contribute to a globally diverse, complex, and interconnected world. This includes creating an inclusive and multicultural educational landscape through the retention and recruitment of diverse students in terms of their backgrounds, identities and experiences, who have been traditionally underrepresented in graduate education. The program promotes a climate of diversity, inclusion, engagement, and achievement, which are integral components of graduate education and beyond.

Student organizations

The Data Science Club at Indiana University (DSC@IU) is a student-run organization affiliated with SICE. All MS Data Science Residential students are encouraged to actively participate in the club. DSC@IU helps students acquire vital skills that will kick-start their journey into the Data Science world, through various means like *mentorships*, *tutorials*, *seminars* and *study groups*. The Club organizes networking meetups for students to connect with Alumni, Professionals, and Employers for career guidance.

Moreover, it conducts Hackathons and Datathons to get hands-on experience with real-world problems and brings great opportunities to socialize through *fun events*. For information about the Data Science Club, email dsclub@indiana.edu.

- Join Facebook at https://www.facebook.com/dsciu/
- Join Slack at https://join.slack.com/t/dsciub/signup
- Join Data Science Club at IU https://beinvolved.indiana.edu/organization/dsciu

Prerequisites and Remedial Coursework

Students who are admitted to the Master of Data Science degree are thought to be ready to start the program with the essential knowledge to be successful in the program. They are not required to take remedial coursework.

However, if a student feels they need remedial work in math and/or programming, they may want to consider enrolling in the Data Science Essentials remedial self-paced package of online coursework that can help you prepare to be successful in the program. The remedial courses available in the Data Science Essentials are: Basic Linear Algebra & Calculus; Basics of Java; Basics of Python Programming; Introduction to C++, Introduction to R Programing; Introduction to SQL; and Introduction to MongoDB. No certificates or badges of understanding will be awarded as these are self-paced modules. This course is offered through the Luddy Office of Online Education (luddyonl@indiana.edu). The cost of this course is $150.

Curriculum

The Master of Data Science degree is a **30-credit degree program** offered by the Luddy School of Informatics, Computing, and Engineering. The curriculum consists of 15 credits of core requirements and the remaining 15 credits are satisfied by fulfilling the Master of Data Science track-specific requirements and electives.

Internship Course Credits

Curricular Practical Training (CPT) and DSCI-D591 Graduate Internship Course (0-3 credits)

The CPT opportunity enables students to work for a company or organizations in the U.S. as an integral part of the established Data Science education program. Students apply their academic knowledge of the in-demand technical skills, working well in a group, interpreting data findings effectively to an audience in various formats, as well as the soft skills employers seek, prepare our graduates to use their expertise in the industry. Each internship will vary according to the context, industry, responsibilities, and personal experiences of the student.

Whether an internship is required for degree completion?

No, CPT is highly recommended but not required for the Data Science degree.

Who is eligible for an internship?

A full-time F-1 student in the Data Science Graduate program with permission from the department.

What is the workflow of CPT?

Here is the information page from OIS at Indiana University and SEVIS information from the US Department of Homeland Security. Students can fill out the CPT form and work with the Luddy Graduate Studies offices and OIS to get an approval. Note that the student will be awarded 1 credit for part-time (1-20 hours) and 2 credits for full-time (more than 21 hours). All internships provide opportunities to build skills in communication, time management and team dynamics. The student can reinforce their academic learnings with the understanding of a real-world corporate setting, which will help the student to make important decisions about a future career in Data Science.

Do I need to enroll in DSCI-D591?

Yes, CPT is an experience to be integral to the student’s academic curriculum. The DSCI-D591 course is being created to allow students the opportunity to gain professional work experience in industry and to utilize skills taught in the classroom. At the completion of the internship, an exit letter from the employer commending the intern’s performance and a written report from the student on the experience will be submitted and the Faculty will evaluate the written report and give the student grades for DSCI-D591.

2. Applied Data Science Track Overview and Track Requirements

The Applied Data Science track offers the training in both the data science methods and their application in different domains. This track is suitable for students with an interdisciplinary background who want to specialize in application areas of data science.

Students following the Applied Data Science track are required to complete **12** credit hours of core coursework that covers **3** credit hours of Statistical Methods, **3** credit hours of Data Mining and Search, **3** credit hours of Data Management and Engineering, and **3** credit hours of Data Visualization and Storytelling. Students will specialize in **6** credit hours of a Data Science Domain. The remaining **12** credit hours are **3** credit hours of capstone project and **9** credit hours of electives, selected to best suit individual interests, needs, and overall career goals.

Applied Data Science Core Requirements (12 credit Hours)

Statistical Methods (3 credit hours)

- **STAT-S 520 Introduction to Statistics (3 cr.)** P: MATH M212, M301, M303, or the equivalent. Basic concepts of data analysis and statistical inference, applied to 1-sample and 2-sample location problems, the analysis of variance, and linear regression. Prob-ability models and statistical methods applied to practical situ-ations and actual data sets from various disciplines. Elementary statistical theory, including the plug-in principle, maximum likelihood, and the method of least squares.
    - Higher level statistics course may be taken with departmental approval

Data Mining and Search (3 credit hours)

- **CSCI-B 551 Elements of Artificial Intelligence (3 cr.)** CSCI-C 343 recommended. Introduction to major issues and approaches in artificial intelligence. Principles of reactive, goal-based, and utility-based agents. Problem-solving and search. Knowledge representation and design of representational vocabularies. Inference and theorem proving, reasoning under uncertainty, and planning. Overview of machine learning.
- **CSCI-B 555 Machine Learning (3 cr.)** Theory and practice of constructing algorithms that learn functions and choose optimal decisions from data and knowledge. Topics include: mathematical/probabilistic foundations, MAP classification/regression, linear and logistic regression, neural networks, support vector machines, Bayesian networks, tree models, committee machines, kernel functions, EM, density estimation, accuracy estimation, normalization, model selection.
- **CSCI-B 565 Data Mining (3 cr.)** Algorithmic and practical aspects of discovering patterns and relationships in large databases. The course also provides hands-on experience in data analysis, clustering and prediction. Topics include data preprocessing and exploration, data warehousing, association rule mining, classification and regression, clustering, anomaly detection, human factors and social issues in data mining.
- **CSCI-P 556 Applied Machine Learning (3 cr.)** The main aim of the course is to provide skills to apply machine learning algorithms on real applications. We will consider fewer learning algorithms and less time on math and theory and instead spend more time on hands-on skills required for algorithms to work on a variety of data sets.
- **ENGR-E 511 Machine Learning for Signal Processing (3 cr.)** The course discusses advanced signal processing topics as an application of machine learning. Hands-on signal processing tasks are introduced and tackled using a problem-solving manner, so students can grasp important

machine learning concepts. The course can help students learn to build an intelligent signal processing system in a systematic way.

- **ILS-Z 534 Search (3 cr.)** The success of commercial search engines shows that Information Retrieval is a key in helping users find the information they seek. This course provides an introduction to information retrieval theories and concepts underlying all search applications. We investigate techniques used in modern search engines and demonstrate their significance by experiment.
- **INFO-I 606 Network Science (3 cr.)** (may be counted only once) Requires strong working knowledge of mathematics and programming, specifically, proficiency in the topics such as probability, statistics, linear algebra, data structures, and algorithms. Python is the main programming language. This course teaches the fundamental theories, algorithms, and key applications of network science across social and biological systems.

Data Management and Engineering (3 credit hours)

Select one course from the following:

- **CSCI-B 561 Advanced Database Concepts (3 cr.)** CSCI-C 241, C 335 and C 343 recommended. Database models and systems: especially relational and object-oriented; relational database design theory; structures for efficient data access; query languages and processing; database applications development; views. Transaction management: concurrency and recovery. Credit not given for both CSCI-B 561 and CSCI-B 461.
- **ENGR-E 516 Engineering Cloud Computing (3 cr.)** Experience with Windows or Linux using Java and scripts. This course covers basic concepts on programming models and tools of cloud computing to support data intensive science applications. Students will get to know the latest research topics of cloud platforms, parallel algorithms, storage, and high-level language for proficiency with a complex ecosystem of tools that span many disciplines.
- **INFO-I 535 Management, Access, and Use of Big and Complex Data (3 cr.)** Innovation today is emerging from a preponderance of data from sensors, social media, and the Internet. This course covers knowledge representation, data process, and data management for big and complex data. Specific topics include data integration, semantics, and provenance; workflows and pipelines; and distributed noSQL stores. Credit not given for both INFO-I 535 and I 435.

Data Visualization and Storytelling (3 credit hours)

Select one course from the following:

- **ENGR-E 583 Information Visualization (3 cr.)** (may only be counted once.) This course provides students with a working knowledge on how to visualize abstract information and hands-on experience in the application of this knowledge to specific domains, different tasks, and diverse, possibly non-technical users. Credit not given for both ENGR-E 583 and E 483.
- **ENGR-E 584 Scientific Visualization (3 cr.)** Teaches basic principles of human cognition and perception; techniques and algorithms for designing and critiquing scientific visualizations in different domains (neuro, nano, bio-medicine, IoT, smart cities); hands-on experience using modern tools for designing scientific visualizations that provide novel and/or actionable insights; 3D printing and augmented reality deployment; teamwork/project management expertise.
- **INFO-I 590 Topics in Informatics**
    - **Topic: Data Visualization (3 cr.)** (may be counted only once) From dashboards in a car to cutting-edge scientific papers, we extensively use visual representation of data. As our world becomes increasingly connected and digitized and as more decisions are being driven by data, data visualization is becoming a critical skill for every knowledge worker. In this course we will learn fundamentals of data visualization and create visualizations that can provide insights into complex datasets.
- **STAT-S 670 Exploratory Data Analysis (3 cr.)** P: Two statistics courses at the graduate level, or consent of instructor. Numerical and graphical techniques for summarizing and displaying data.

Ex-ploration versus confirmation. Connections with conventional statistical analysis and data mining. Applications to large data sets.

Applied Data Science Domain (6 credit hours)

Select one of the following domains and complete two courses within that specific domain:

Augmented and Virtual Reality

- **INFO-I 590 Topics in Informatics**
    - **Topic: Artificial Life in Virtual Reality (3 cr.)** P: INFO-I 304. This course will explore one powerful application of virtual reality: the study of life, evolution, and artificial intelligence. Students will learn the basic building blocks of biological intelligence, how to build virtual worlds for assessing artificial intelligence, and how to populate virtual worlds with intelligent and autonomous artificial agents.
    - **Topic: Building Virtual Worlds (3 cr.)** P: INFO-I 304. This course will explore advanced techniques for designing and building virtual reality worlds. Topics include rigged animation, spatial sound, keyframe and procedural animation, interactivity, intelligent cameras, advanced shaders, and particle systems. Students will develop proficiency with a variety of software tools, development methods, and creation techniques.
    - **Topic: Creating Virtual Assets (3 cr.)** P: INFO-I 304. This course will explore advanced techniques for creating virtual assets for virtual reality applications. Topics include 3D modeling, animation, motion capture, sound capture and editing, materials, textures, shaders, and scripting. Students will learn how to export assets to virtual reality, augmented reality, video, still images, and 3D printed objects.
    - **Topic: Introduction to Virtual Reality (3 cr.)** Virtual Reality has applications in fields as diverse as medicine, education, military training, trauma recovery, and artificial intelligence. In this course, students will learn the foundational skills needed to build virtual reality applications. We will focus on software programs for building virtual assets and realistic virtual environments.

Data Security and Privacy

- **INFO-I 520 Security for Networked Systems (3 cr.)** This course is an extensive survey of system and network security. Course materials cover the threats to information confidentiality, integrity and availability and the defense mechanisms that control such threats. The course provides the foundation for more advanced security courses and hands-on experiences through course projects. Credit not given for both INFO-I 520 and I 430.INFO-I 525 Organizational Informatics and Economics of Security
- **INFO-I 525 Organizational Informatics and Economics of Security (3 cr.)** Security technologies make explicit organizational choices that allocate power. Security implementations allocate risk, determine authority, reify or alter relationships, and determine trust extended to organizational participants. The course begins with an introduction to relevant definitions (security, privacy, trust) and then moves to a series of timely case studies of security technologies.
- **INFO-I 533 Systems and Protocol Security and Information Assurance (3 cr.)** This course looks at systems and protocols, how to design threat models for them and how to use a large number of current security technologies and concepts to block specific vulnerabilities. Students will use a large number of systems and programming security tools in the laboratories. Credit not given for both INFO-I 533 and I 433.
- **INFO-I 538 Introduction to Cryptography (3 cr.)** This class considers issues of network security, treating in depth the topics covered in INFO-I 536. In particular, the class involves adversarial modeling, a detailed treatment of security primitives, and methods for analysis of security. It spans the ethics and technology of security, with examples drawn both from deployed and proposed protocols. Topics to be covered include studies of rational and malicious cheating, symmetric and asymmetric cryptography, security reductions and heuristics.

Health and Biomedical Data Science

- **INFO-I 507 Introduction to Health Informatics (3 cr.)** This is a combined advanced undergraduate and graduate course that provides an introduction to health informatics. By the end of the course, students will be able to describe and apply informatics methods that improve health and well being.
- **INFO-I 519 Introduction to Bioinformatics (3 cr.)** One semester programming course or equivalent recommended. Sequence alignment and assembly; RNA structure, protein and molecular modeling; genomics and proteomics; gene prediction; phylogenetic analysis; information and machine learning; visual and graphical analysis bioinformatics; worldwide biological databases; experimental design and data collection techniques; scientific and statistical data analysis; database and data mining methods; and network and Internet methods.
- **INFO-I 529 Machine Learning in Bioinformatics (3 cr.)** INFO-I 519 or equivalent knowledge recommended. The course covers advanced topics in Bioinformatics with a focus on machine learning. The course will review existing techniques such as hidden Markov models, artificial neural networks, decision trees, stochastic grammars, and kernel methods. Examine application of these techniques to current bioinformatics problems including: genome annotation and comparison, gene finding, RNA secondary structure prediction, protein structure prediction, gene expression analysis, proteomics, and integrative functional genomics.

Human Robotic Interaction

- **CSCI-B 657 Computer Vision (3 cr.)** P: CSCI-B 551. Concepts and methods of machine vision as a branch of artificial intelligence. Basics of digital image processing. Local and global tools for deriving information from image data. Model-based object recognition and scene understanding.
- **ENGR-E 523 Internet of Things (3 cr.)** Java, C, and Python will be used as programming languages. This course covers the Internet of Things (IoT) including the emerging Industrial IoT. Power, security, networking, system architecture from cloud to device are covered. Integration with big data and use cases are discussed. Laboratory sessions are integrated.
- **ENGR-E 599 Topics in Intelligent Systems Engineering**
  - **Topic: Autonomous Robotics (3 cr.)**
- **INFO-I 513 Usable Artificial Intelligence (3 cr.)** Building foundational skills in machine learning, natural language processing, and artificial intelligence for data collection, data analysis, data visualization, and decision-making.
- **INFO-I 527 Mobile and Pervasive Design (3 cr.)** The aim of this course is to provide students with the ability to design and implement novel interactions with mobile and pervasive technologies. We will discuss interaction paradigms and explore different technologies. Students will design, build, implement and refine mobile and pervasive computing applications for their domain of interest.
- **INFO-I 540 Human Robot Interaction (3 cr.)** This course surveys the field of human-robot interaction (HRI), which involves understanding how people perceive and respond to robots and creating robots that interact naturally with people. We will discuss the design, evaluation and societal significance of interactive robots from a human-centered perspective through readings, discussion and developing HRI prototypes. Credit given for only one of INFO-I 540, I 440 or H 440.
- **INFO-I 542 Foundations of HCI (3 cr.)** "Foundations of HCI" offers a survey overview of the field of Human-Computer Interaction Design. It introduces the main themes of HCI set generally in a historical context. Themes include interaction design, cognitive modeling, distributed cognition, computer-supported cooperative work, data visualization, ubiquitous computing, affective computing, and domestic computing, among others.

Social Data Science

- **ENGR-E 583 Information Visualization (3 cr.)** (may only be counted once.) This course provides students with a working knowledge on how to visualize abstract information and hands-

on experience in the application of this knowledge to specific domains, different tasks, and diverse, possibly non-technical users. Credit not given for both ENGR-E 583 and E 483.
- **ILS-Z 639 Social Media Mining (3 cr.)** Basic Unix skills recommended. This course provides a graduate-level introduction to social media mining and methods. The course provides hands-on experience mining social data for social meaning extraction (focus on sentiment analysis) using automated methods and machine learning technologies. We will read, discuss, and critique claims and findings from contemporary research related to SMM.
- **INFO-I 513 Usable Artificial Intelligence (3 cr.)** Building foundational skills in machine learning, natural language processing, and artificial intelligence for data collection, data analysis, data visualization, and decision-making.
- **INFO-I 590 Topics in Informatics**
    - **Topic: Data Visualization (3 cr.)** (may be counted only once) From dashboards in a car to cutting-edge scientific papers, we extensively use visual representation of data. As our world becomes increasingly connected and digitized and as more decisions are being driven by data, data visualization is becoming a critical skill for every knowledge worker. In this course we will learn fundamentals of data visualization and create visualizations that can provide insights into complex datasets.
- **INFO-I 606 Network Science (3 cr.)** (may be counted only once) Requires strong working knowledge of mathematics and programming, specifically, proficiency in the topics such as probability, statistics, linear algebra, data structures, and algorithms. Python is the main programming language. This course teaches the fundamental theories, algorithms, and key applications of network science across social and biological systems.

Applied Data Science Capstone Project (3 credit Hours)

Students will be required to work on a project that applies the knowledge and skills learned to solve real-world problems for a company, organization, or individual. This may be fulfilled through a capstone course, an internship, or an independent study project. The aim of this requirement is to demonstrate students' capabilities to prospective employers and inspire innovation.

- **DSCI-D 591 Graduate Internship (0-3 cr.)** Department Approval. Students gain professional work experience in an industry or research organization setting using skills and knowledge acquired in Data Science coursework. A written report will be required upon completion of the experience. May be repeated for a maximum of 6 credit hours.
- **DSCI-D 592 Data Science in Practice (3 cr.)** Students gain critical, practical skills in applying data science to real world problems. Students will work in teams of 3-5 to tackle a real-world problem defined by a project sponsor. Project sponsors can be academics or industry practitioners. Students work with the project sponsor to understand the problem domain, identify where their data science skills can be applied, and to design, implement and test a solution.
- **DSCI-D 699 Graduate Independent Study in Data Science (1-6 cr.)** Must be a student in the Data Science graduate program. Independent Study under the direction of a faculty member, culminating in a written report and/or database development and/or documented laboratory experience. May be repeated 2 times for a maximum of 9 credit hours.
- **ILS-Z 690 Capstone in Information Architecture (3 cr.)** The capstone course integrates within a single project the theoretical and practical components of the Information Architecture Certificate program. Working with one of the program co-directors, who serves as the student's project advisor, the student will determine both the scope and extent of the project. The student will publicly present and defend the capstone project upon completion.

Applied Data Science Electives (9 credit Hours)

The remaining 9 credit hours are selected from unselected courses above or additional data science-related course offerings within the Luddy School of Informatics, Computing, and Engineering. Students may not earn credit for courses taken to fulfill core, domain, or capstone requirements.

- No more than three (3) credit hours of DSCI-D 591 may be earned

- No more than three (3) credit hours total may be earned in DSCI-D 590 Basic Data Science On-Ramp and DSCI-D 590 Advanced Data Science On-Ramp

Applied Data Science Track Sample Schedule of Courses

The following is a sample schedule of courses for the Applied Data Science Track. Students should consult with their advisor and the Director of Graduate Studies in order to select courses that will best support their plans and career goals.

Sample Schedule of Courses

Fall Year 1 (9 cr.)	**Spring Year 1** (9 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Data Science Domain Course (3 cr.)	Data Science Doman Course (3 cr.)

Fall Year 2 (9 cr.)	**Spring Year 2** (3 cr.)
Elective (3 cr.)	Capstone Project (3 cr.)
Elective (3 cr.)	Elective (3 cr.) (optional)
Elective (3 cr.)	Elective (3 cr.) (optional)

Applied Data Science Track Degree Audit Form

Core Requirement (15 credit hours)			
Statistical Methods (3 cr.)		**Data Management and Engineering (3 cr.)**	
STAT-S 520 Introduction to Statistics • Higher level Statistics course may be taken with approval		ENGR-E 516 Engineering Cloud Computing	
Data Mining and Search (3 cr.)		CSCI-B 561 Advanced Database Concepts	
CSCI-B 551 Elements of AI		INFO-I 535 Management, Access, and Use of Big and Complex Data	
CSCI-B 555 Machine Learning		**Data Visualization and Storytelling (3 cr.)**	
CSCI-B 565 Data Mining		ENGR-E 583 Information Visualization	
CSCI-P 556 Applied Machine Learning		ENGR-E 584 Scientific Visualization	
ENGR-E 511 Machine Learning for Signal Processing		INFO-I 590 Topic: Data Visualization	
ILS-Z 534 Search		STAT-S 670 Exploratory Data Analysis	
INFO-I 606 Network Science			
Data Science Domain (6 credit hours)			
Augmented and Virtual Reality		**Data Security and Privacy**	
INFO-I 590 Topic: Artificial Life in Virtual Reality		INFO-I 520 Security for Networked Systems	
INFO-I 590 Topic: Building Virtual Worlds		INFO-I 525 Organizational Informatics and Economics of Security	
INFO-I 590 Topic: Creating Virtual Assets		INFO-I 533 Systems and Protocol Security and Information Assurance	
INFO-I 590 Topic: Introduction to Virtual Reality		INFO-I 538 Introduction to Cryptography	
Human Robotic Interaction		**Health and Biomedical Data Science**	
CSCI-B 657 Computer Vision		INFO-I 507 Introduction to Health Informatics	
ENGR-E 523 Internet of Things		INFO-I 519 Introduction to Bioinformatics	
ENGR-E 599 Topic: Autonomous Robotics		INFO-I 529 Machine Learning in Bioinformatics	
INFO-I 513 Usable Artificial Intelligence		**Social Data Science (3 cr.)**	
INFO-I 527 Mobile and Pervasive Design		ENGR-E 583 Information Visualization	
INFO-I 540 Human Robot Interaction		ILS-Z 639 Social Media Mining	
INFO-I 542 Foundations of HCI		INFO-I 513 Usable Artificial Intelligence	
		INFO-I 590 Topic: Data Visualization	
		INFO-I 606 Network Science	
Electives (9 credit hours)		**Capstone Project (3 credit hours)**	
		DSCI-D 591 Graduate Internship	
		DSCI-D 592 Data Science in Practice	
		DSCI-D 699 Independent Study in Data Science	
		ILS-Z 690 Capstone in Information Architecture	

3. Big Data Systems Track Overview and Requirements

The Big Data Systems track focuses on the development and engineering of software systems for collecting, managing, and mining massive data. This is most suitable for students with a background in computer science or engineering who prefer hands-on and project-based learning.

Students following the Big Data Systems track are required to complete **21** credit hours of core coursework that covers **3** credit hours of Statistical Methods, **6** credit hours of AI and Machine Learning, **9** credit hours of Big Data, Cloud Computing, and Visualization, and **3** credit hours of Core Engineering. The remaining **9** credit hours are electives, selected to best suit individual interests, needs, and overall career goals.

Pre-requisites: Students in this program need to have a solid foundation in STEM course work, specifically the following:

- Proficient level of programming experience in C, Java or Python
- Familiarity with R and MATLAB is useful
- Calculus I and II and basic understanding of probability and elements of discrete math

Big Data Systems Track Core Requirements (21 credit Hours)

Statistical Methods (3 credit hours)

Select one course from the following:

- SPEA-V 506 Statistical Analysis for Effective Decision-making
- STAT-S 520 Introduction to Statistics
    - Students who have completed equivalent prior coursework in statistics can opt to take an additional elective in lieu of one of the Statistical Methods courses

AI and Machine Learning for Engineering (6 credit hours)

Select two courses from the following:

- **CSCI-B 555 Machine Learning (3 cr.)** Theory and practice of constructing algorithms that learn functions and choose optimal decisions from data and knowledge. Topics include: mathematical/probabilistic foundations, MAP classification/regression, linear and logistic regression, neural networks, support vector machines, Bayesian networks, tree models, committee machines, kernel functions, EM, density estimation, accuracy estimation, normalization, model selection.
- **CSCI-B 565 Data Mining (3 cr.)** Algorithmic and practical aspects of discovering patterns and relationships in large databases. The course also provides hands-on experience in data analysis, clustering and prediction. Topics include data preprocessing and exploration, data warehousing, association rule mining, classification and regression, clustering, anomaly detection, human factors and social issues in data mining.
- **CSCI-P 556 Applied Machine Learning (3 cr.)** The main aim of the course is to provide skills to apply machine learning algorithms on real applications. We will consider fewer learning algorithms and less time on math and theory and instead spend more time on hands-on skills required for algorithms to work on a variety of data sets.

- **ENGR-E 511 Machine Learning for Signal Processing (3 cr.)** The course discusses advanced signal processing topics as an application of machine learning. Hands-on signal processing tasks are introduced and tackled using a problem-solving manner, so students can grasp important machine learning concepts. The course can help students learn to build an intelligent signal processing system in a systematic way.
- **ENGR-E 533 Deep Learning Systems (3 cr.)** This course teaches the pipeline for building state-of-the-art deep learning-based intelligent systems. It covers general training mechanisms and acceleration options that use GPU computing libraries and parallelization techniques running on high performance computing systems. The course also aims at deploying the networks to the low-powered hardware systems.
- **ENGR-E 635 Big Data Graph Analytics (3 cr.)** This course covers theoretical and practical concepts in large-scale graph analytics with applications to social networks, computational biology, machine learning and scientific computing. It will demonstrate graph algorithms by analyzing large-scale social and biological networks using high-performance graph analytics frameworks. Design principles for parallel graph algorithms will be discussed.

Big Data, Cloud Computing, and Visualization (9 credit hours)

Select three courses from the following:

- **CSCI-B 561 Advanced Database Concepts (3 cr.)** CSCI-C 241, C 335 and C 343 recommended. Database models and systems: especially relational and object-oriented; relational database design theory; structures for efficient data access; query languages and processing; database applications development; views. Transaction management: concurrency and recovery. Credit not given for both CSCI-B 561 and CSCI-B 461.
- **ENGR-E 516 Engineering Cloud Computing (3 cr.)** Experience with Windows or Linux using Java and scripts. This course covers basic concepts on programming models and tools of cloud computing to support data intensive science applications. Students will get to know the latest research topics of cloud platforms, parallel algorithms, storage, and high-level language for proficiency with a complex ecosystem of tools that span many disciplines.
- **ENGR-E 522 HPC and Cloud Computing for Large Scale Image Applications (3 cr.)** Java and Python will be used as programming languages. Understanding of machine learning and/or image processing is helpful. This course describes big data techniques for sensors and remote sensing explaining how one architects analysis systems for sensors and remote imagery. Algorithms, software systems, and storage issues are addressed. The impact of user interfaces is covered. Streaming and batch examples from satellite, internet of things and physics data.
- **ENGR-E 534 Big Data Applications (3 cr.)** This is an overview course of Big Data Applications covering a broad range of problems and solutions. It covers cloud computing technologies and includes a project. Algorithms are introduced and illustrated.
- **ENGR-E 583 Information Visualization (3 cr.)** This course provides students with a working knowledge on how to visualize abstract information and hands-on experience in the application of this knowledge to specific domains, different tasks, and diverse, possibly non-technical users. Credit not given for both ENGR-E 583 and E 483.
- **ENGR-E 584 Scientific Visualization (3 cr.)** Teaches basic principles of human cognition and perception; techniques and algorithms for designing and critiquing scientific visualizations in different domains (neuro, nano, bio-medicine, IoT, smart cities); hands-on experience using modern tools for designing scientific visualizations that provide novel and/or actionable insights; 3D printing and augmented reality deployment; teamwork/project management expertise.
- **ENGR-E 616 Advanced Cloud Computing (3 cr.)** This course describes Cloud 3.0 in which DevOps, Microservices, and Function as a Service is added to basic cloud computing. The discussion is centered around the Apache Big Data Stack and a major student project aimed at demonstrating integration of cloud capabilities.
- **ENGR-E 623 Applied Streaming Data Systems** Java, C, and Python will be used as programming languages. This course covers the software and algorithm engineering of streaming data systems in the cloud with an emphasis on use in industry and the internet of things.

Core Engineering (3 credit hours)

Select one course from the following:

- **ENGR-E 503 Introduction to Intelligent Systems (3 cr.)** This course covers fundamental principles and five use cases with special attention to challenges and opportunities coming from modern computing infrastructure, the internet of things and artificial intelligence.
- **ENGR-E 517 High Performance Computing (3 cr.)** Students will learn about the development, operation, and application of HPC systems prepared to address future challenges demanding capability and expertise. The course combines critical elements from hardware technology and architecture, system software and tools, and programming models and application algorithms with the cross-cutting theme of performance management and measurement.
- **ENGR-E 523 Internet of Things (3 cr.)** This course covers the Internet of Things (IoT) including the emerging Industrial IoT. Power, security, networking, system architecture from cloud to device are covered. Integration with big data and use cases are discussed. Laboratory sessions are integrated.
- **ENGR-E 535 Image Processing for Medical Applications (3 cr.)** Learn how to build intelligent algorithms and software for medical imaging that can help medical doctors to treat their patients and researchers to understand how the body works. Students will be familiarized with algorithmic techniques such as tracking, denoising, warping, segmentation, model fitting, optimization and interactive visualization of medical datasets.
- **ENGR-E 551 Simulating Nanoscale Systems (3 cr.)** Students will learn how to model and simulate material behavior at the nanoscale. Analysis and control of shape, assembly, and flow behavior in soft nanomaterials will be discussed. Applications to engineering problems at the nanoscale will be emphasized. Optimization methods, nonequilibrium systems, and parallel computing will be covered.

Big Data Systems Electives (9 credit Hours)

The remaining 9 credit hours can be selected from unselected courses above or additional data science-related course offerings within the Luddy School of Informatics, Computing, and Engineering. Students may not earn credit for courses taken to fulfill core requirements.

- No more than three (3) credit hours of DSCI-D 591 may be earned
- No more than three (3) credit hours total may be earned in DSCI-D 590 Basic Data Science On-Ramp and DSCI-D 590 Advanced Data Science On-Ramp

Big Data Systems Track Sample Schedule of Courses

The following is a sample schedule of courses for the Big Data Systems Track. Students should consult with their advisor and the Director of Graduate Studies in order to select courses that will best support their plans and career goals.

Sample Schedule of Courses

Fall Year 1 (9 cr.)	**Spring Year 1** (9 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Elective (3 cr.)

Fall Year 2 (9 cr.)	**Spring Year 2** (3 cr.)
Core Course (3 cr.)	Elective (3 cr.)
Core Course (3 cr.)	Elective (3 cr.) (optional)
Elective (3 cr.)	Elective (3 cr.) (optional)

Big Data Systems Track Degree Audit Form

Core Requirement (21 credit hours)			
Statistical Methods (3 cr.)		**AI and Machine Learning for Engineering (6 cr.)**	
SPEA-V 506 Statistical Analysis for Effective Decision-making		CSCI-B 555 Machine Learning	
STAT-S 520 Introduction to Statistics • Students who have completed equivalent prior coursework in statistics can opt to take an additional elective in lieu of one of the Statistical Methods courses		CSCI-B 565 Data Mining	
Big Data, Cloud Computing, and Visualization (3 cr.)		CSCI-B 561 Advanced Database Concepts	
CSCI-B 561 Advanced Database Concepts		ENGR-E 511 Machine Learning for Signal Processing	
ENGR-E 516 Engineering Cloud Computing		ENGR-E 533 Deep Learning Systems	
ENGR-E 522 HPC and Cloud Computing for Large Scale Image Applications		ENGR-E 635 Big Data Graph Analytics	
ENGR-E 534 Big Data Applications		**Core Engineering (3 cr.)**	
ENGR-E 583 Information Visualization		ENGR-E 503 Introduction to Intelligent Systems	
ENGR-E 584 Scientific Visualization		ENGR-E 517 High Performance Computing	
ENGR-E 616 Advanced Cloud Computing		ENGR-E 523 Internet of Things	
ENGR-E 623 Applied Streaming Data Systems		ENGR-E 535 Image Processing for Medical Applications	
		ENGR-E 551 Simulating Nanoscale Systems	
Electives (9 credit hours)			

4. Computational and Analytical Track Overview and Requirements

The Computational and Analytical track focuses on the foundational data science methods. This track is most suitable for students with a background in computer science, statistics, or mathematics who wish to drive deeper into the mechanics of data science methodologies.

Students following the Computational and Analytical track are required to complete **15** credit hours of core coursework that covers **3** credit hours of Data Systems Foundation, **3** credit hours of Algorithmic Foundation, **6** credit hours of Data Analytics Foundation, and **3** credit hours of Big Data Infrastructures. The remaining **15** credit hours are electives, selected to best suit individual interests, needs, and overall career goals.

Computational and Analytical Core Requirements (15 credit Hours)

Data Systems Foundation (3 credit hours)

- **CSCI-B 561 Advanced Database Concepts (3 cr.)** CSCI-C 241, C 335 and C 343 recommended. Database models and systems: especially relational and object-oriented; relational database design theory; structures for efficient data access; query languages and processing; database applications development; views. Transaction management: concurrency and recovery. Credit not given for both CSCI-B 561 and B 461.

Algorithmic Foundation (3 credit hours)

Select one course from the following:

- **CSCI-B 503 Algorithms Design and Analysis (3 cr.)** Models, algorithms, recurrences, summations, growth rates. Probabilistic tools, upper and lower bounds; worst-case and average-case analysis, amortized analysis, dynamization. Comparison-based algorithms: search, selection, sorting, hashing. Information extraction algorithms (graphs, databases). Graphs algorithms: spanning trees, shortest paths, connectivity, depth-first search, breadth-first search.
- **CSCI-B 505 Applied Algorithms (3 cr.)** The course studies the design, implementation, and analysis of algorithms and data structures as applied to real world problems. The topics include divide-and-conquer, optimization, and randomized algorithms applied to problems such as sorting, searching, and graph analysis. The course teaches trees, hash tables, heaps, and graphs.

Data Analytics Foundations (6 credit hours)

- **STAT-S 520 Introduction to Statistics (3 cr.)** P: MATH M212, M301, M303, or the equivalent. Basic concepts of data analysis and statistical inference, applied to 1-sample and 2-sample location problems, the analysis of variance, and linear regression. Prob-ability models and statistical methods applied to practical situ-ations and actual data sets from various disciplines. Elementary statistical theory, including the plug-in principle, maximum likelihood, and the method of least squares.
    - Higher level statistics course may be taken with departmental approval

Select one additional course from the following:

- **CSCI-B 555 Machine Learning (3 cr.)** Theory and practice of constructing algorithms that learn functions and choose optimal decisions from data and knowledge. Topics include: mathematical/probabilistic foundations, MAP classification/regression, linear and logistic regression, neural networks, support vector machines, Bayesian networks, tree models, committee machines, kernel functions, EM, density estimation, accuracy estimation, normalization, model selection.
- **CSCI-B 565 Data Mining (3 cr.)** Algorithmic and practical aspects of discovering patterns and relationships in large databases. The course also provides hands-on experience in data analysis, clustering and prediction. Topics include data preprocessing and exploration, data warehousing, association rule mining, classification and regression, clustering, anomaly detection, human factors and social issues in data mining.

Big Data Infrastructure (3 credit hours)

Select one course from the following:

- **ENGR-E 516 Engineering Cloud Computing (3 cr.)** Experience with Windows or Linux using Java and scripts. This course covers basic concepts on programming models and tools of cloud computing to support data intensive science applications. Students will get to know the latest research topics of cloud platforms, parallel algorithms, storage, and high-level language for proficiency with a complex ecosystem of tools that span many disciplines.
- **INFO-I 535 Management, Access, and Use of Big and Complex Data (3 cr.)** Innovation today is emerging from a preponderance of data from sensors, social media, and the Internet. This course covers knowledge representation, data process, and data management for big and complex data. Specific topics include data integration, semantics, and provenance; workflows and pipelines; and distributed noSQL stores. Credit not given for both INFO-I 535 and I 435.

Computational and Analytical Electives (15 credit Hours)

The remaining 15 credit hours can be selected from unselected courses above or additional data science-related course offerings within the Luddy School of Informatics, Computing, and Engineering. Students may not earn credit for courses taken to fulfill core requirements.

- No more than three (3) credit hours of DSCI-D 591 may be earned
- No more than three (3) credit hours total may be earned in DSCI-D 590 Basic Data Science On-Ramp and DSCI-D 590 Advanced Data Science On-Ramp

Computational and Analytical Track Sample Schedule of courses

The following is a sample schedule of courses for the Computational and Analytical Track. Students should consult with their advisor and the Director of Graduate Studies in order to select courses that will best support their plans and career goals.

Sample Schedule of Courses

Fall Year 1 (9 cr.)	**Spring Year 1** (9 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Elective (3 cr.)	Elective (3 cr.)

Fall Year 2 (9 cr.)	**Spring Year 2** (3 cr.)
Core Course (3 cr.)	Elective (3 cr.)
Elective (3 cr.)	Elective (3 cr.) (optional)
Elective (3 cr.)	Elective (3 cr.) (optional)

Computational and Analytical Track Degree Audit Form

Core Requirement (15 credit hours)			
Data Systems Foundation (3 cr.)		**Algorithmic Foundation (3 cr.)**	
CSCI-B 561 Advanced Database Concepts		CSCI-B 503 Algorithms Design and Analysis	
Data Analytics Foundation (6 cr.)		CSCI-B505 Applied Algorithms	
STAT-S 520 Introduction to Statistics • Higher level Statistics course may be taken with approval		**Big Data Infrastructure (3 cr.)**	
CSCI-B 555 Machine Learning		ENGR-E 516 Engineering Cloud Computing	
CSCI-B 565 Data Mining		INFO-I 535 Management, Access, and Use of Big Complex Data	
Electives (15 credit hours)			

5. Managerial Data Science Track Overview and Requirements

The managerial data science track combines advanced knowledge in database systems and programming languages with strong interpersonal and project management skills. This track is most suitable for students with prior work experience who wish to develop organizational and project management skills.

Students following the Managerial Data Science track are required to complete **21** credit hours of core coursework that covers **3** credit hours of Statistical Methods, **3** credit hours of Machine Learning, Data Mining, and Text Mining, **3** credit hours of Data Visualization and Storytelling, **6** credit hours of Management in Theory, and **6** credit hours of Management in Practice. The remaining **9** credit hours are **3** credit hours of capstone project and **6** credit hours of electives, selected to best suit individual interests, needs, and overall career goals.

Statistical Methods (3 credit hours)

Select one course from the following:

- **SPEA-V 506: Statistical Analysis for Effective Decision-Making (3 cr.)** Non-calculus survey of concepts in probability, estimation, and hypothesis testing. Applications of contingency table analysis; analysis of variance, regression, and other statistical techniques. Computer processing of data emphasized.
- **STAT-S 520: Introduction to Statistics (3 cr.)** Basic concepts of data analysis and statistical inference, applied to 1-sample and 2-sample location problems, the analysis of variance, and linear regression. Probability models and statistical methods applied to practical situations and actual data sets from various disciplines. Elementary statistical theory, including the plug-in principle, maximum likelihood, and the method of least squares.
    - Higher level statistics course may be taken with departmental approval

Machine Learning, Data Mining, and Text Mining (3 credit hours)

Select one course from the following:

- **CSCI-B 505 Applied Algorithms (3 cr.)** The course studies the design, implementation, and analysis of algorithms and data structures as applied to real world problems. The topics include divide-and-conquer, optimization, and randomized algorithms applied to problems such as sorting, searching, and graph analysis. The course teaches trees, hash tables, heaps, and graphs.
- **CSCI-B 551 Elements of Artificial Intelligence (3 cr.)** CSCI-C 343 recommended. Introduction to major issues and approaches in artificial intelligence. Principles of reactive, goal-based, and utility-based agents. Problem-solving and search. Knowledge representation and design of representational vocabularies. Inference and theorem proving, reasoning under uncertainty, and planning. Overview of machine learning.
- **CSCI-B 555 Machine Learning (3 cr.)** Theory and practice of constructing algorithms that learn functions and choose optimal decisions from data and knowledge. Topics include: mathematical/probabilistic foundations, MAP classification/regression, linear and logistic regression, neural networks, support vector machines, Bayesian networks, tree models, committee machines, kernel functions, EM, density estimation, accuracy estimation, normalization, model selection.

- **CSCI-B 561 Advanced Database Concepts (3 cr.)** CSCI-C 241, C 335 and C 343 recommended. Database models and systems: especially relational and object-oriented; relational database design theory; structures for efficient data access; query languages and processing; database applications development; views. Transaction management: concurrency and recovery. Credit not given for both CSCI-B 561 and CSCI-B 461.
- **CSCI-B 565 Data Mining (3 cr.)** Algorithmic and practical aspects of discovering patterns and relationships in large databases. The course also provides hands-on experience in data analysis, clustering and prediction. Topics include data preprocessing and exploration, data warehousing, association rule mining, classification and regression, clustering, anomaly detection, human factors and social issues in data mining.
- **CSCI-B 657 Computer Vision (3 cr.)** Concepts and methods of machine vision as a branch of artificial intelligence. Basics of digital image processing. Local and global tools for deriving information from image data. Model-based object recognition and scene understanding.
- **CSCI-P 556 Applied Machine Learning (3 cr.)** The main aim of the course is to provide skills to apply machine learning algorithms on real applications. We will consider fewer learning algorithms and less time on math and theory and instead spend more time on hands-on skills required for algorithms to work on a variety of data sets.
- **ENGR-E 511 Machine Learning for Signal Processing (3 cr.)** The course discusses advanced signal processing topics as an application of machine learning. Hands-on signal processing tasks are introduced and tackled using a problem-solving manner, so students can grasp important machine learning concepts. The course can help students learn to build an intelligent signal processing system in a systematic way.
- **ILS-Z 534 Search (3 cr.)** The success of commercial search engines shows that Information Retrieval is a key in helping users find the information they seek. This course provides an introduction to information retrieval theories and concepts underlying all search applications. We investigate techniques used in modern search engines and demonstrate their significance by experiment.
- **INFO-I 513 Usable Artificial Intelligence (3 cr.)** Building foundational skills in machine learning, natural language processing, and artificial intelligence for data collection, data analysis, data visualization, and decision-making.
- **INFO-I 606 Network Science (3 cr.)** (may be counted only once) Requires strong working knowledge of mathematics and programming, specifically, proficiency in the topics such as probability, statistics, linear algebra, data structures, and algorithms. Python is the main programming language. This course teaches the fundamental theories, algorithms, and key applications of network science across social and biological systems.

Data Visualization and Storytelling (3 credit hours)

Select one course from the following:

- **ENGR-E 583 Information Visualization (3 cr.)** This course provides students with a working knowledge on how to visualize abstract information and hands-on experience in the application of this knowledge to specific domains, different tasks, and diverse, possibly non-technical users. Credit not given for both ENGR-E 583 and E 483.
- **ENGR-E 584 Scientific Visualization (3 cr.)** Teaches basic principles of human cognition and perception; techniques and algorithms for designing and critiquing scientific visualizations in different domains (neuro, nano, bio-medicine, IoT, smart cities); hands-on experience using modern tools for designing scientific visualizations that provide novel and/or actionable insights; 3D printing and augmented reality deployment; teamwork/project management expertise.
- **INFO-I 590 Topics in Informatics**
    - **Topic: Data Visualization (3 cr.)** (may be counted only once) From dashboards in a car to cutting-edge scientific papers, we extensively use visual representation of data. As our world becomes increasingly connected and digitized and as more decisions are being driven by data, data visualization is becoming a critical skill for every knowledge worker. In this course we will learn fundamentals of data visualization and create visualizations that can provide insights into complex datasets.

Management in Theory (6 credit hours)

Select two courses from the following:

- **ILS-Z 513 Organizational Informatics (3 cr.)** Introduction to information, technology, and social behavior in the organizational context. Concepts of organization theory and organization behavior, including knowledge and information management, organizational analytics, and organizational intelligence, provide a critical foundation for managing information, people, and information and communication technologies (ICTs) in rapidly changing and dynamic environments.
- **ILS-Z 645 Social and Organizational Informatics of Big Data (3 cr.)** This course surveys organizational, legal, political, and social issues surrounding the creation, dissemination and use of big data from the perspective of social and organizational informatics. It focuses on ways in which the integration of big data is changing structure, culture, and work practices in private and public sector organizations.
- **ILS-Z 604 Topics in Library and Information Science**
    - **Topic: Data Ethics (3 cr.)**

Management in Practice (6 credit hours)

Select two courses from the following:

- **ILS-Z 512: Information Systems Design (3 cr.)** Students identify, design, and implement a significant information design project, such as the redesign of a complex Website for a local business, library, or nonprofit. Principles and practices of project management are discussed in the context of team-based website redesign.
- **ILS-Z 556: Systems Analysis & Design (3 cr.)** This course introduces the basic concepts underlying systems analysis and design, focusing on contextual inquiry/design and data modeling, as well as the application of those analysis techniques in the analysis and design of organizational information systems.
- **ILS-Z 586 Digital Curation (3 cr.)** Preserving and providing long-term access to digital materials over time is a Grand Challenge. They require constant and ongoing maintenance. This course provides an overview of research, policy and current practices in curating and preserving digital data, gives students practical experience, working with digital materials, and creating digital curation plans.

Managerial Data Science Capstone Project (3 credit Hours)

Students will be required to work on a project that applies the knowledge and skills learned to solve real-world problems for a company, organization, or individual. This may be fulfilled through a capstone course, an internship, or an independent study project. The aim of this requirement is to demonstrate students' capabilities to prospective employers and inspire innovation.

- **DSCI-D 591 Graduate Internship (0-3 cr.)** Department Approval. Students gain professional work experience in an industry or research organization setting using skills and knowledge acquired in Data Science coursework. A written report will be required upon completion of the experience. May be repeated for a maximum of 6 credit hours.
- **DSCI-D 592 Data Science in Practice (3 cr.)** Students gain critical, practical skills in applying data science to real world problems. Students will work in teams of 3-5 to tackle a real-world problem defined by a project sponsor. Project sponsors can be academics or industry practitioners. Students work with the project sponsor to understand the problem domain, identify where their data science skills can be applied, and to design, implement and test a solution.
- **DSCI-D 699 Graduate Independent Study in Data Science (1-6 cr.)** Must be a student in the Data Science graduate program. Independent Study under the direction of a faculty member, culminating in a written report and/or database development and/or documented laboratory experience. May be repeated 2 times for a maximum of 9 credit hours.

- **ILS-Z 690 Capstone in Information Architecture (3 cr.)** The capstone course integrates within a single project the theoretical and practical components of the Information Architecture Certificate program. Working with one of the program co-directors, who serves as the student's project advisor, the student will determine both the scope and extent of the project. The student will publicly present and defend the capstone project upon completion.

Managerial Data Science Electives (6 credit Hours)

The remaining 6 credit hours can be selected from unselected courses above or additional data science-related course offerings within the Luddy School of Informatics, Computing, and Engineering. Students may not earn credit for courses taken to fulfill core requirements.

- No more than three (3) credit hours of DSCI-D 591 may be earned
- No more than three (3) credit hours total may be earned in DSCI-D 590 Basic Data Science On-Ramp and DSCI-D 590 Advanced Data Science On-Ramp

Managerial Data Science Track Sample Schedule of Courses

The following is a sample schedule of courses for the Managerial Data Science Track. Students should consult with their advisor and the Director of Graduate Studies in order to select courses that will best support their plans and career goals.

Sample Schedule of Courses

Fall Year 1 (9 cr.)	**Spring Year 1** (9 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Core Course (3 cr.)
Core Course (3 cr.)	Elective (3 cr.)

Fall Year 2 (9 cr.)	**Spring Year 2** (3 cr.)
Core Course (3 cr.)	Capstone Project (3 cr.)
Core Course (3 cr.)	Elective (3 cr.) (optional)
Elective (3 cr.)	Elective (3 cr.) (optional)

Managerial Data Science Track Course Checklist form

Core Requirement (21 credit hours)			
Statistical Methods (3 cr.)		**Machine Learning, Data Mining, and Text Mining (3 cr.)**	
SPEA-V 506 Statistical Analysis for Effective Decision-making		CSCI-B 505 Applied Algorithm	
STAT-S 520 Introduction to Statistics • Students who have completed equivalent prior coursework in statistics can opt to take an additional elective in lieu of one of the Statistical Methods courses		CSCI-B 551 Elements of Artificial Intelligence	
Data Visualization and Storytelling (3 cr.)		CSCI-B 555 Machine Learning	
ENGR-E 583 Information Visualization		CSCI-B 561 Advanced Database Concepts	
ENGR-E 584 Scientific Visualization		CSCI-B 565 Data Mining	
INFO-I 590 Topic: Data Visualization		CSCI-B 567 Computer Vision	
Management in Theory (6 cr.)		CSCI-P 556 Applied Machine Learning	
ILS-Z 513 Organizational Informatics		ENGR-E 511 Machine Learning for Signal Processing	
ILS-Z 645 Social and Organizational Informatics of Big Data		ILS-Z 534 Search	
ILS-Z 604 Data Ethics		INFO-I 513 Usable Artificial Intelligence	
Management in Practice (6 cr.)		INFO-I 606 Network Science	
ILS-Z 512 Information Systems Design			
ILS-Z 556 Systems Analysis and Design			
ILS-Z 586 Digital Curation			
Electives (6 credit hours)		**Capstone Project (3 credit hours)**	
		DSCI-D 591 Graduate Internship	
		DSCI-D 592 Data Science in Practice	
		DSCI-D 699 Independent Study in Data Science	
		ILS-Z 690 Capstone in Information Architecture	

6. Other Resources

Course Substitution and Exceptions

Any course substitution or exception must be approved in advance. Email the Luddy Graduate Studies Office for additional information (gradvise@indiana.edu).

Grades

The minimum overall GPA of a grade of B (3.0) for all Ph.D. Informatics courses is required. A student whose semester GPA falls below a grade of B (3.0) will be put on probation. The student must raise their semester and cumulative grade point average to a B (3.0) or higher by the end of the following semester. Failure to do so in two consecutive semesters (excluding summer) may result in academic dismissal from the program.

Grade Appeals

If a student believes there has been an error in calculating the final grade in a course, the student may appeal that grade. For information on the grade appeal process, please follow the instructions provided by the Student Advocates Office at https://studentaffairs.indiana.edu/student-support/advocates/help/academic-help/changes-appeals.html.

Academic Probation

A student will be placed on academic probation if the student's cumulative or semester GPA falls below a 3.0 and/or if a student fails to make satisfactory progress in the program. To return to satisfactory progress status, students must bring their cumulative and semester grade point averages to 3.0 or higher by the end of the next semester. Failure to do so may result in academic dismissal from the program.

Registering for Classes

To help with the registration process, students are given a Course Planning Checklist. Students meet with their advisor prior to registering to plan courses for the upcoming semester. The student may email gradvise@indiana.edu to schedule an appointment.

Some courses require course permission from the instructor and/or the department prior to enrollment. This information is found in the Schedule of Classes which is located at http://registrar.indiana.edu/calendars/schedule-of-classes.shtml. If the course is listed as requiring permission from the instructor or the department, students must contact via email the instructor and/or the department listed for the course to obtain permission. The email reply must be forwarded to gradvise@indiana.edu.

Independent study classes and all research classes taken prior to entering candidacy require that the student and the instructor define the study/rotation, including the deliverables. Students should complete the Independent Study/Rotation/Research Agreement, obtain the signed permission of the instructor supervising the study/rotation, and submit it to gradvise@indiana.edu along with the Course Registration Form.

After all approvals are secured, students should complete the Course Registration Form and ask their advisor to sign it. Students should then send registration and agreement forms to gradvise@indiana.edu.

The Luddy Graduate Studies Office will process the form and notify students by email of any issues or that students may proceed with registration for the term.

Students then register for courses via one.iu.edu.

Instructions for how to register are found at websites for Student Central https://studentcentral.indiana.edu/register/steps-register/index.html and the Enrollment and Student Academic Information Bulletin http://enrollmentbulletin.indiana.edu/pages/registration.php?t=fall#procedure

How to Register for courses and enrollment at the shopping center?

To register for classes, a student will need their IU network ID username, passphrase, and DUO to log into one.iu.edu.

How to Register for Classes and Enrollment Shopping Cart (https://kb.iu.edu/d/anig)

- **Determining whether students have holds on their record** (https://kb.iu.edu/d/anig#holds)
- **Viewing class permissions** (https://kb.iu.edu/d/anig#perm)
- **Using the Enrollment Shopping Cart** (https://kb.iu.edu/d/anig#cart)
  - Adding classes (https://kb.iu.edu/d/anig#adding)
  - Registering from the shopping cart (https://kb.iu.edu/d/anig#regcart)
- **Using Class Registration** (https://kb.iu.edu/d/anig#regdrop)
  - Registering for classes (https://kb.iu.edu/d/anig#regclass)
  - Dropping a class (https://kb.iu.edu/d/anig#dropclass)
  - Editing classes with variable credit (https://kb.iu.edu/d/anig#variable)
  - Swapping classes (https://kb.iu.edu/d/anig#swapping)
- **Viewing class schedule details** (https://kb.iu.edu/d/anig#det)

Additional steps on how to register are available through the UITS Knowledge Base: http://www.kb.iu.edu/data/anig.html.

Class Permissions

Any deviations from a student's approved Course Registration Form requires that the student request approval from the advisor or from the program director if the course to be dropped/added is a program core course or otherwise required. Approval should then be conveyed in writing (email or signed document) to the Luddy Graduate Studies Office, gradvise@indiana.edu.

***Fees/Refund.** Starting two business days after the student's initial registration, a system access fee of $8.60 is charged every calendar day the student makes one or more successful adjustments to their schedule. A $23 late schedule change fee is assessed for each course dropped after the first week of classes. The late schedule change fee also applies to a section change, a change of arranged hours, or an audit change.

Students are responsible for paying all drop and add fees. 100% of tuition is refunded for a course dropped during the first week of classes. After the first week, the amount of tuition refunded (if any) for a dropped course depends on the type of session the course is and when the course is dropped.

*Fees are current at the time of publication and are subject to change.

Waitlist

If a course is shown as full, the student should add themselves to the waitlist, which serves as a placeholder in the registration line. When students who enrolled in the course drop or when the enrollment cap is expand-ed, students on the waitlist will be admitted into the course in order. Note: The waitlist runs for the last time on the Thursday of the first week of classes. The Drop if Enroll feature allows a student to enroll in another course while waitlisted for their course of first preference. Students must remember to cancel this feature if they decide to remain in the class of their second choice. The Swap feature allows a student to delay dropping a course until they are safely enrolled in their new class.

Class Withdrawal

During the automatic withdrawal period (see the Registrar's Official Calendar for exact dates), students who withdraw will be assigned an automatic grade of W. After that period, withdrawals are only possible with approval from the Dean, which is normally given only for urgent reasons such as illness. Instructors may award a grade of F for a student who is failing and withdraws after the automatic withdrawal period.

Bursar Bill

Tuition, fees, and all other charges (e.g., IU Health Center, IU Library) are billed to the student on their Bursar bill. Payments are due the 10th of the month. For a list of the Bursar Bill Due Dates go to https://studentcentral.indiana.edu/pay-for-college/pay-bill/due-dates.html.

To pay the Bursar bill, students may pay in person at the Poplars Building from 9:00 a.m. to 4:30 p.m. on weekdays (except holidays). See the alternative payment methods at Payment Options, https://studentcentral.indiana.edu/pay-for-college/pay-bill/payment-options/index.html.

Graduation

All graduate students are encouraged to participate in Commencement. Indiana University hosts two university wide commencement events – Winter and Spring. The majority of the students attend the Spring Commencement. Students who finish their degree during the fall can attend the Winter or Spring Commencement. The solemn yet colorful academic pageantry can provide a fitting culmination to a period of intense study and work.

Visit https://commencement.indiana.edu/index.html for detailed information. Be sure to watch for these emails as many of the deadlines are time sensitive.

In addition to Indiana University's Commencement Event, the Luddy School of Informatics, Computing, and Engineering hosts a Celebration Event. Be sure to watch for these emails as many of the deadlines are time sensitive.

Diplomas

Diplomas will read "Master of Science in Data Science."

Transcripts

Transcripts will read "Master of Data Science."

Funding

All completed applications that are reviewed by the admissions committee are considered for financial awards which Informatics may be offering. If a student is admitted, the student's admit letter will detail the financial award, if any.

Every year, the school has limited funds to distribute. Depending on the number of applications received and the competitiveness of those applications determines the number of financial awards.

Funding, if awarded, may take the form of a (1) 10-hour-per-week appointment for duties within the Luddy School of Informatics, Computing, and Engineering, (2) fellowship award, and (3) travel award.

If students are awarded a 10-hour-per-week appointment, students are required to fulfill their appointment responsibilities of grading finals and other administrative duties through the end of finals week for both the fall and spring semesters. Please refer to IU's official academic calendar for official dates. Failure to fulfill appointment responsibilities may result in termination of the appointment.

Full-time Status and Part-time Status

A student must be enrolled in a minimum of eight (8) credit hours each semester to be considered full-time. Audited courses are not counted in the definition of "full-time study." It is imperative that international students maintain full-time status to remain in visa compliance. For questions about visa compliance, contact the Office of International Services (ois@iu.edu).

Approval must be given for a student to be enrolled as a part-time student (less than 8 credit hours). Email the Luddy Graduate Studies Office for additional information (gradvise@indiana.edu).

Leave of Absence

A leave of absence allows Informatics graduate students to deal with unforeseen events that interfere with their academic progress. During a leave, the student is not expected to make progress toward the degree. Although the student may complete coursework from previous terms during a leave, the student may not attend class or use the leave to catch up on current coursework, prepare for exams, work on the capstone, and/or the master's thesis project.

To be eligible for a leave, the student must be enrolled full time in an Informatics graduate program and have completed at least one semester (a minimum of nine credit hours) in the program. The student must be in good academic standing—if they are on academic probation, they are not eligible for a leave.

Approval must be given for a student to take a leave of absence (less than 8 credit hours). Email the Luddy Graduate Studies Office for additional information (gradvise@indiana.edu).

Important Student Resources

Throughout this handbook, you will see references to the following important websites, forms, and other resources:

- ACM Code of Ethics and Professional Conduct
- Indiana University Code of Student Rights, Responsibilities, and Conduct
- Indiana University Graduate School Bulletin
- Degree Programs
- Graduate Program Courses
- Indiana Graduate Student Academic Appointees Guide
- Indiana University Office of International Services (Email: ois@iu.edu)
- one.iu.edu
- University Graduate School (UGS)

Luddy Graduate Studies Office Contact

Luddy Graduate Student Office (GSO)
Luddy Hall, Room 1113
700 N Woodlawn Ave
Bloomington, Indiana 47408

Advising: gradvise@indiana.edu

Admissions: admit2iu@indiana.edu

Luddy Graduate Studies Office (GSO) is committed to supporting graduate students from admissions to graduation. Luddy GSO can assist with questions about:

- Graduate Admissions
- New Student Orientation
- Academic Graduate Policies
- Program Administrative Forms
- Transfer Credit Requests
- Grade Changes
- Degree Audits
- Ph.D. interested in getting MS/MA
- Under Enrollment Requests
- Probation Issues
- Registration (late/add/drop/withdraw)
- Doctoral Milestones
- Ph.D. Annual Evaluations – also known as GED
- Leave of Absence
- Advising Holds
- Program Transfer within SICE
- Student Travel
- A11 Holds
- R10 Immunizations
- V03 Academic Hold
- Fee Waiver Requests
- International Student Services
    - CPT/OPT e-form requests
    - Academic Training Requests
    - Under enrollment permission
    - I-20 Extensions