DATA SCIENCE, MASTER OF SCIENCE - DSCI

Program Description

A Data Scientist is a professional who combines many types of technical and industry competencies to turn data, which is very often idiosyncratic and ambiguous, into actionable intelligence in a business environment. The skills needed to make this transformation draw from mathematics, statistics, computer science, business, and require the ability to communicate technical information to people with a range of technical competence. The Master of Science in Data Science is a rigorous program designed to rapidly bring students to the point of functioning in the role of a data scientist and then, building upon the initial growth, to develop expertise with their data science skills.

The program in Data Science has several components. It requires coursework over a two-year period in mathematics, statistics, and computer science that supports the program outcomes. The program is centered on core data science courses including an introduction to data mining and applied data analytics. Supporting courses include applied statistics, applied linear algebra, computer programming, and databases. It also requires coursework that uses core knowledge and skills in a professional environment, such as technical communication and professional writing. The program includes a capstone project that provides a substantive professional context for students to apply their data science knowledge.

Prerequisites and Core Competencies

The most competitive candidates will satisfy the prerequisites and core competencies as follows:

- A quantitative undergraduate major (examples include but are not limited to mathematics, the sciences, social sciences, and business with a quantitative emphasis) or a career in a technical, or quantitative area
- One semester of calculus preferred
- Familiarity with computer programming
- Familiarity with statistics

Candidates who meet some but not all of the prerequisites and core competencies are encouraged to apply and will be considered conditionally. The Program Director can identify opportunities for those candidates to gain familiarity in the relevant area(s).

Application Requirements

- A bachelor's degree from a regionally accredited school, or the international equivalent.
- Candidates should have core competencies which may be demonstrated by her/his education : a bachelor's degree in mathematics, business, computer science, information systems, the sciences, health science, quantitative social science or related field; the most competitive candidates will have at least a 3.0 cumulative GPA in undergraduate coursework.
- Submission of a completed application including the following:
  - Official transcripts from your degree-granting institutions.
  - Personal statement that describes how the experiences in your life make you ideally suited to become a data scientist.

Applications open in September for entry into the program the following fall. The application deadline is rolling, and applications will be accepted as long as seats are available in the entering class.

4+1/3+1 Pathway for Saint Mary's Undergraduate Students

Saint Mary's College students meeting the prerequisites set forth below may apply to the Master of Science in Data Science program starting in the second semester sophomore year. If admitted to the graduate program, students will complete two to four graduate courses in Data Science in the junior and senior year prior to baccalaureate graduation. The student will continue Data Science courses in the summer term immediately following her baccalaureate graduation and continue for the next fall, spring, and summer terms to complete the bachelors and graduate degree in five years. In the fifth year, students will be charged the per credit hour rate equivalent to the cohort she is joining.

Prerequisites:

- Calculus I
- Statistics
- Computer Programming

Practicum Presentation

All students are required to give a formal presentation about the project completed for the DSCI 599 Practicum.

Program in Data Science

Master of Science in Data Science (30-33 hours)

Code	Title	Credits
CPSC 507	Computer Programming	3
CPSC 529	Database Systems	3
DSCI 501	Data Mining	3
DSCI 502	Advanced Topics in Data Science	3
DSCI 511	Data Preprocess/Visualization	3
MATH 527	Linear Algebra for Data Science	2
MATH 548	Statistical Methods for Data Science	3
COMM 503	Communication and Data Science	3
ENWR 517	Professional & Tech Writing	3
PHIL 557	Data Ethics	1
DSCI 599	Practicum	1-6
Total Credits		**28-33**

Student Learning Outcomes

The Master of Science in Data Science program is committed to providing graduates with the range and depth of expertise to be leaders in data driven industries. Students who successfully complete the program will demonstrate high levels of mathematical, analytical, technical, and professional skills and knowledge. Upon the completion of the program, students will be able to:

- Analyze large, complex data sets as would be encountered in the context of real-world business problems.

- Apply and fine-tunes computing resources for data analysis, including programming and industry-standard tool use.
- Develop and implements data analysis strategies based on theoretical principles, ethical considerations, and detailed knowledge of the underlying data.
- Generate actionable intelligence for decision-making.
- Clearly and professionally communicates nuanced analysis results to a diverse, varyingly-technical audience.
- Rigorously apply mathematical principles to the analysis of data.
- Evaluate, implement, and assess the application of technology solutions for data analysis.
- Plan, direct, and evaluate the status of complex projects.

Program Director

Bogdan Vajiac
48 Madeleva Hall
574-284-4717

Faculty

C. Fitzpatrick, C. Hoover, J. Juszkiewicz, K. Kuter, E. Misiolek, A. Pierce, R. Rohatgi, N. Salehi, B. Vajiac, C. Wedrychowicz,

Data Science CourseS

COMM 503 Communication and Data Science (3)
Industry experts stress the importance of often-overlooked communication skills in data science. Rachel Hawley, Analytic Solutions Architect at the SAS Institute, states "it is extremely important that potential candidates have effective communication and presentation skills. It's not enough to just have the technical chops, a data scientist must be able to effectively explain how he or she came to a specific conclusion and convince the internal or external customer that their results should be leveraged." This course is designed to explore this intersection between communication and data science. Topics will include assessing and improving communication skills, interpersonal and intercultural communication, teamwork, and leadership. The development of effective presentational skills, particularly oral skills, will be stressed.

CPSC 507 Computer Programming (3)
A problem-solving approach to learning computer programming. Topics include variables, data types, conditional statements, loops, arrays, recursion, principles of software engineering, object-oriented programming, data structures, algorithms, and the use of standard libraries available in a variety of programming languages. The course will use commerically common programming languages and integrated development environments (IDEs).

CPSC 529 Database Systems (3)
Basic concepts of databases. Topics include conceptual data modeling, database design and normalization, and database implementation. Use of SQL for data definition, manipulation, and query processing. While primary emphasis will be on the relational model and traditional RDBMS, discussion will also include a survey of techniques for handling non-relational data models, massive datasets, and unstuctured data, including data warehousing, in-memory databases, NewSQL, NoSQL and Hadoop.

DSCI 501 Data Mining (3)
This course is about mining knowledge from data in order to gain useful insights and predictions. From theory to practice, the course investigates all stages of the knowledge discovery process, which includes data preprocessing, exploratory data analysis, prediction and discovery through regression and classification, clustering, association analysis, anomaly detection, and postprocessing.

DSCI 502 Advanced Topics in Data Science (3)
Advanced Topics in Data Science is a comprehensive course designed to provide students with both foundational and advanced concepts in data science and machine learning. The course begins with an introduction to programming and data analysis using Python, equipping students with essential coding and analytical skills. It covers core machine learning techniques, including regression methods, classification approaches, and anomaly detection, before advancing into deep learning with neural networks (NNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs). Students will also explore dimensionality reduction with Principal Component Analysis (PCA) and learn about autoencoders for unsupervised representation learning. A key focus of the course is the practical application of these techniques, culminating in model deployment using Amazon SageMaker. This hands-on approach prepares students to develop and implement scalable machine learning solutions in real-world environments. Prerequisite:DSCI 501

DSCI 511 Data Preprocess/Visualization (3)
This course is an introduction to data visualization. It includes data preprocessing and focuses on specific tools and techniques necessary to visualize complex data. Data visualization topics covered include design principles, perception, color, statistical graphs, maps, trees and networks, and other topics as appropriate. Visualization tools may include JavaScript D3 library, Python, and R, and commercially available software such as Tableau, etc. The course introduces the techniques necessary to successfully implement visualization projects using the programming languages studied.

DSCI 595 Thesis (1-3)
Thesis credit may be earned for significant work toward the writing of a master's thesis. This thesis may be used to fulfill the culminating project requirement.

DSCI 599 Practicum (1-6)
The practicum is an opportunity to directly experience the work of a data scientist or data analytics professional. It consists of project-based learning on a significant and contributory business objective in conjunction with practicing professionals in one of many appropriate industries. May be repeated up to 6 credits.

ENWR 517 Professional & Tech Writing (3)
This course teaches skills in written, visual, and verbal communication of particular importance to data science professionals. It engages with foundational concepts of rhetoric, composition, and design that can be applied in any setting while also addressing the forms and conventions of technical writing in a professional setting that students will encounter as practicing researchers and data analysts. The course stresses the seamless continuity between analysis of data and communication about that analysis.

MATH 527 Linear Algebra for Data Science (2)
An application-focused approach to linear algebra used in data science. Topics include matrices, Gaussian elimination, vector spaces, inner products, orthogonality, least squares, eigenvalues/vectors, matrix factorizations, singular value decomposition and principal component analysis, quadratic forms, data/image processing, and other topics pertinent to data analysis.

MATH 548 Statistical Methods for Data Science (3)
This course provides a comprehensive, application-focused overview of essential statistical methods for data science. Topics include data collection techniques, descriptive statistics, and exploratory data analysis. Foundational concepts such as sampling distributions and the Central Limit Theorem set the groundwork for estimation, confidence intervals, and hypothesis testing. Students will explore techniques in ANOVA and categorical data analysis, as well as nonparametric techniques and permutation tests. Advanced methods include the bootstrap, linear and logistic regression, generalized linear models, and linear discriminant analysis, equipping students with a versatile toolkit for real-world data analysis and decision-making in data-driven contexts.

NURS 670 Nursing Informatics & Data-Driven Decision Making (3)
This course is designed to equip DNP students with essential knowledge and skills in nursing informatics, data analytics, and their application to improve healthcare outcomes. This course emphasizes the significant role of nursing informatics in today's evolving healthcare landscape and explores the opportunities and challenges of integrating informatics and data analytics in various healthcare settings. Students will gain an understanding of the use of electronic health records, telehealth, and clinical decision support systems in enhancing patient care and safety. Students will develop competencies in using informatics tools and techniques to analyze large volumes of data, supporting evidence-based nursing practice. Additionally, the course will examine ethical and legal considerations and advocacy related to informatics and data analytics in nursing practice. Prerequisites: NURS 612; NURS 620; NURS 622.

PHIL 557 Data Ethics (1)
Data about us is collected continuously, and in many ways makes our lives as we know them possible—enabling your doctors to treat you efficiently, letting Amazon show you what you need to buy before you even know it exists, helping Spotify introduce you to the next music you'll love. But is there a dark side to all this data-driven convenience? In this one credit hour course, students will engage with the ethical challenges posed by data collection, analysis and use, through class discussion, case study analysis and course readings. We begin by considering various ethical frameworks, including utilitarianism and deontology. We then engage with the history of data collection, looking at the abuse of humans, particularly from marginalized groups, in the Nazi experiments, the Tuskegee syphilis experiments and the history of eugenics in the US. Turning to contemporary methods of collecting and using data, we consider key areas of ethical concern including: issues of autonomy and consent, privacy and surveillance, artificial intelligence and machine learning, disinformation and bias, and algorithmic discrimination. Students demonstrate mastery of the material in online discussion, brief writing assignments, and analysis of a self-chosen contemporary case study.

Degree Plans in Data Science

The academic plans for the 3+1 and the 4+1 require two summer courses (one between years 3 and 4 and one after year 4), and students would be taking graduate courses over two years.

- Sample Schedule for a 3+1 student in Mathematics and Data Science (p. 3)
- Sample Schedule for a 3+1 student in Computing and Applied Mathematics and Data Science (p. 4)
- Sample Schedule for a 4+1 student in Data Science (p. 5)

Sample Schedule for a 3+1 student in Mathematics and Data Science

Course	Title	Credits
First Year		
First Semester		
MATH 131	Calculus I	4
AVE 101	College in Practice	1
Language I		3
ENLT W		4
Natural Science		3
First Year Seminar		3
	Credits	**18**
Second Semester		
CPSC 207	Computer Programming	3
MATH 132	Calculus II	4
Language II		3
PHIL course		3
Elective Class		3
	Credits	**16**
Second Year		
First Semester		
MATH 225	Foundations of Higher Mathematics	3
MATH 231	Calculus III	4
RLST1 course		3
Social Science course		3
ART course		3
Elective Class		3
	Credits	**19**
Second Semester		
MATH 326	Linear Algebra and Differential Equations	4
MATH 345	Probability	3
RLST2 course		3
Interdisciplinary Thinking course		3
HIST course		3
Elective Class		3
	Credits	**19**
Third Year		
First Semester		
MATH 346	Statistics	3
MATH 353	Abstract Algebra I	3
MATH 527	Linear Algebra for Data Science	2
CPSC 507	Computer Programming	3
Elective Class		3
	Credits	**14**
Second Semester		
MATH 354	Abstract Algebra II	3
MATH 339 or MATH 372	Discrete Mathematics (or MATH 3XX elective) or Stochastic Models	3
MATH 548	Statistical Methods for Data Science	3
DSCI 511	Data Preprocess/Visualization	3

Elective Class		3
	Credits	**15**
Third Semester		
ENWR 517	Professional & Tech Writing	3
	Credits	**3**
Fourth Year		
First Semester		
MATH 341	Analysis I	3
MATH 496	Pro-Seminar	2
DSCI 501	Data Mining	3
DSCI 599	Practicum	3
Elective Class		3
	Credits	**14**
Second Semester		
DSCI 502	Advanced Topics in Data Science	3
DSCI 599	Practicum	3
Elective Class		3
Elective Class		3
Elective Class		3
Graduate Undergraduate		
	Credits	**15**
Third Semester		
Graduate Student		
COMM 503	Communication and Data Science	3
Graduate MS in Data Science		
	Credits	**3**
	Total Credits	**136**

Sample Schedule for a 3+1 student in Computing and Applied Mathematics and Data Science

Course	Title	Credits
First Year		
First Semester		
MATH 131	Calculus I	4
AVE 101	College in Practice	1
Language I		3
ENLT W course		4
Natural Science course		3
First Year Seminar course		3
	Credits	**18**
Second Semester		
CPSC 207	Computer Programming	3
MATH 132	Calculus II	4
Language II course		3
PHIL course		3
Elective Class		3
	Credits	**16**
Second Year		
First Semester		
MATH 225	Foundations of Higher Mathematics	3
MATH 231	Calculus III	4
RLST 1		3
Social Science course		3
ART course		3
Elective Class		3
	Credits	**19**
Second Semester		
MATH 326	Linear Algebra and Differential Equations	4
MATH 345	Probability	3
RLST 2 course		3
Interdisciplinary Thinking course		3
HIST course		3
CPSC 328	Data Structures	3
	Credits	**19**
Third Year		
First Semester		
CPSC 417	Systems Analysis and Design	4
MATH 381	Mathematical Modeling	3
MATH 527	Linear Algebra for Data Science	2
CPSC 507	Computer Programming	3
Elective Class		3
	Credits	**15**
Second Semester		
CPSC 308	Electronic Communications	3
MATH 339	Discrete Mathematics	3
MATH 548	Statistical Methods for Data Science	3
DSCI 511	Data Preprocess/Visualization	3
Elective Class		3
	Credits	**15**
Third Semester		
ENWR 517	Professional & Tech Writing	3
	Credits	**3**
Fourth Year		
First Semester		
CPSC 429	Database Systems	3
MATH 496	Pro-Seminar	2
DSCI 501	Data Mining	3
DSCI 599	Practicum	3
Elective Class		3
	Credits	**14**
Second Semester		
DSCI 502	Advanced Topics in Data Science	3
DSCI 599	Practicum	3
Elective Class		3
Elective Class		3
Graduate Undergraduate		
	Credits	**12**
Third Semester		
Graduate Student		
COMM 503	Communication and Data Science	3
Graduate MS in Data Science		
	Credits	**3**
	Total Credits	**134**

Sample Schedule for a 4+1 student in Data Science

4+1 Schedule as Undergraduate

Course	Title	Credits
Third Year		
First Semester		
CPSC 507	Computer Programming	3
Any Junior class may be moved to Senior.		
CPSC 529 is an optional class for students who already completed CPSC 429.		
	Credits	**3**
Second Semester		
MATH 548	Statistical Methods for Data Science	3
	Credits	**3**
Third Semester		
ENWR 517	Professional & Tech Writing	3
	Credits	**3**
Fourth Year		
First Semester		
MATH 527	Linear Algebra for Data Science	2
	Credits	**2**
Second Semester		
DSCI 511	Data Preprocess/Visualization	3
Graduate Undergraduate		
	Credits	**3**
	Total Credits	**14**

4+1 Schedule as Graduate Student

Course	Title	Credits
First Year		
First Semester		
Graduate Student - Summer after UG graduation		
COMM 503	Communication and Data Science	3
	Credits	**3**
Second Semester		
DSCI 501	Data Mining	3
DSCI 599	Practicum	3
DSCI 599 is the Purdue Data Mine class (2 semesters). The program requires 30-33 credits.		
	Credits	**6**
Third Semester		
DSCI 502	Advanced Topics in Data Science	3
DSCI 599	Practicum	3
CPSC 529	Database Systems	3
Graduate MS in Data Science		
	Credits	**9**
	Total Credits	**18**