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 communication, professional writing, research methods, and project management. 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 perferred
- Familiarity with computer programming
- Familiarity with statistics
- Familiarity with linear algebra

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 education or experience:
  - **Education:** 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.
  - **Experience:** relevant work experience in a technical or quantitative area.
- Submission of a completed application including the following:
  - Official transcripts from your degree-granting institutions.
  - Current résumé or Curriculum Vitae.
  - One letter of recommendation from academic and/or other professionals addressing your ability to succeed in the program (three recommended).
  - Personal statement that describes how the experiences in your life make you ideally suited to become a data scientist.
  - English language proficiency if your education was in a language other than English (to be shown through results from the TOEFL, IELTS, or completion of the appropriate level in the Saint Mary's College English Language School).
  - A video interview (optional).

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. The early action deadline is January 15. The priority application deadline is February 15.

4+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 as a second semester junior or first semester senior. If admitted to the graduate program, students will complete two graduate courses in Data Science in the 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
- Calculus II
- Statistics
- Computer Programming
- Linear Algebra (or equivalent)
- Completion of one graduate Data Science course (only for those applying in fall semester of her senior year)

Summer Immersion

All students are required to participate in a summer immersion experience on campus the week following the summer term. The summer immersion is an intensive experience during which students work in teams and consult on data science related projects from regional businesses and non-profit organizations.

Practicum Presentation

All students are required to give a formal presentation about the project completed for the DSCI 599 Practicum. The presentation shall be given during the summer orientation/symposium in August following the enrollment in DSCI 599 Practicum.

Program in Data Science

Master of Science in Data Science (36 hours)

Code	Title	Credits
CPSC 507	Computer Programming	3
CPSC 529	Database Systems	3
DSCI 501	Data Mining	3
DSCI 502	Data Mining II	3
DSCI 511	Data Preprocess/Visualization	3
MATH 527	Linear Algebra for Data Science	3
MATH 546	Applied Statistics I	3
MATH 547	Applied Statistics II	3
Six credits of the following:		6
BUAD 546	Project Management	
COMM 503	Communication and Data Science	
DSCI 500A	Essential Calculus for Data Science	
DSCI 500B	Essential Probability Theory for Data Science	
DSCI 525	Research Methods	
ENWR 517	Professional & Tech Writing	
NURS 670	Nursing Informatics & Data-Driven Decision Making	
PHIL 557	Data Ethics	
DSCI 599	Practicum (at least 3 credits)	3-6
Additional graduate credits to total 36 credit hours		0-3
Total Credits		**33-39**

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, D. Mallot, E. Misiolek, R. Rohatgi, S. Rohr, R. Rohatgi, B. Vajiac, C. Wedrychowicz, M. Zwart

Data Science CourseS

BUAD 546 Project Management (3)
The course develops the competencies and skills for planning and controlling projects and understanding interpersonal issues that drive successful project outcomes. Focusing on the introduction of new products and processes, it examines the project management life cycle, defining project parameters, matrix management challenges effective project management tools and techniques, and the role of a project manager.

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 500A Essential Calculus for Data Science (1)
This course introduces the concepts from differential, integral, and multivariate calculus essential for the study of data science. Elements of linear algebra, such as vectors, planes, and matrices, also included. Emphasis on computation and application.

DSCI 500B Essential Probability Theory for Data Science (1)
This course introduces concepts from probability theory essential for the study of data science. Topics include probability spaces, Bayes’ Theorem, random variables, discrete and continuous distributions, specifically the normal distribution, and the Central Limit Theorem. Emphasis on computation and application.

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 Data Mining II (3)
A second semester of data mining introducing tools and techniques necessary for mining large scale data sources. 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 525 Research Methods (3)
An introduction to basic scientific and statistical research methods when dealing with measurements of human and corporate activity. Students read and evaluate current research and translate their ideas into viable research projects. Topics include scholarly writing and presentation, descriptive research methods, quasi-experimental and experimental design, ethical issues, and analytical methods.

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 546 Applied Statistics I (3)
An introduction to the foundations and applications of statistics. Topics include basic concepts of data collection sampling and experimental design, descriptive analysis and graphical displays of data, probability concepts and expectations, normal and binomial distributions, sampling distributions and the Central Limit Theorem, confidence intervals and hypothesis testing, likelihood-based statistics, ANOVA, correlation and simple linear regression.

MATH 547 Applied Statistics II (3)
An application-focused approach to regression analysis and related techniques. Topics include simple and multiple linear regression, weighted and generalized least squares estimators, polynomial regression, exponential regression, model selection, categorical variables and ANOVA, logistic regression, principal component analysis, time series analysis, and other applications of statistics as relevant. Prerequisite: MATH 546

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.