DATA SCIENCE

SPS Certificate website: https://sps.northwestern.edu/program-search/certificates.php

Analytics and Modeling, Graduate Certificate
The ability to harness data for actionable insight is increasingly essential for nearly every sector of the economy. We are awash in data, yet companies and organizations don’t always know how best to leverage their data to meet strategic goals, improve outcomes, or simply gain deeper understanding of their operations. This stackable graduate certificate in analytics and modeling focuses on foundational skills and knowledge for those working in or hoping to work in data science and analytics in any industry.

Students will learn various techniques to test real-world predictions about data. They explore data, identify appropriate models, and validate those models. With a focus on traditional methods of applied statistics, this certificate prepares data scientists to utilize algorithms for predictive modeling and analytics, developing models for marketing, finance, and other business applications.

Analytics Management, Graduate Certificate
The ability to harness data for actionable insight is increasingly essential for nearly every sector of the economy. We are awash in data, yet companies and organizations don’t always know how best to leverage their data to meet strategic goals, improve outcomes, or simply gain deeper understanding of their operations. This stackable graduate certificate in analytics management focuses on foundational skills and knowledge for those working in or hoping to work in data science and analytics in any industry.

Analytics managers bridge the work of analysts and modelers with business operations and strategy to help address future business needs, identify business opportunities, and translate the work of data scientists into language that business management understands. This certificate equips data scientists with the communication and management strategies needed to be data-driven leaders who utilize models, analyses, and statistical data to improve business performance.

Data Engineering, Graduate Certificate
The ability to harness data for actionable insight is increasingly essential for nearly every sector of the economy. We are awash in data, yet companies and organizations don’t always know how best to leverage their data to meet strategic goals, improve outcomes, or simply gain deeper understanding of their operations. This stackable graduate certificate in data engineering focuses on foundational skills and knowledge for those working in or hoping to work in data science and analytics in any industry.

Data engineers implement models to scale within an information infrastructure, creating systems and workflows to organize and manage large quantities of data. This means understanding computer systems and solving problems related to data collection, security, and organization. This certificate trains data scientists to utilize system-wide problem-solving skills, choose hardware systems, and build software systems for implementing models made by data analysts to scale in productions systems. This certificate would appeal to students with strong undergraduate backgrounds in data science, computer science, mathematics, engineering, and information systems programs at Northwestern and elsewhere.

Technology Entrepreneurship, Graduate Certificate
Entrepreneurship involves creating a new business or business function where one did not exist before. Advances in science and technology spur innovation, giving existing, resource-rich companies a chance to reinvent themselves, often moving into new markets. These advances, many of them emerging from computer science, engineering, and data science, provide an opportunity for individuals and firms to build new organizations or startups. This certificate is designed to serve students from many disciplines. It shows the path to building a successful, innovation-driven startup. The certificate will be most useful to students who are thinking about starting new technology companies but have little previous training or experience in business.

Artificial Intelligence, Advanced Graduate Certificate
Recent advances in and integrations of machine learning and artificial intelligence are significantly impacting the ways that data scientists build and program computer software, statistical models, and neural networks. Each of these areas is comprised of different analytical approaches; artificial intelligence utilizes rules-based, logic, and knowledge-based systems, while machine learning relies upon data-adaptive methods such as naïve Bayes models, nearest neighbor models, classification and regression trees, random forests, support vector machines, and neural networks. Data-adaptive models are the next step in the progression of modeling and analytics, from traditional statistics to data mining to machine learning, where computer software learns from data and makes traditional models more flexible. Many of today’s prominent applications of artificial intelligence – including computer vision, natural language processing, and robotics – rely on deep learning, the area where machine learning and artificial intelligence overlap.

Certificates Offered
• Analytics and Modeling, Graduate Certificate (https://catalogs.northwestern.edu/sps/certificates/graduate/data-science/analytics-modeling-graduate-certificate/)
• Analytics Management, Graduate Certificate (https://catalogs.northwestern.edu/sps/certificates/graduate/data-science/analytics-management-graduate-certificate/)
• Artificial Intelligence, Advanced Graduate Certificate (https://catalogs.northwestern.edu/sps/certificates/graduate/data-science/artificial-intelligence-advance-graduate-certificate/)
• Data Engineering, Graduate Certificate (https://catalogs.northwestern.edu/sps/certificates/graduate/data-science/data-engineering-graduate-certificate/)
• Technology Entrepreneurship, Graduate Certificate (https://catalogs.northwestern.edu/sps/certificates/graduate/data-science/technology-entrepreneurship-graduate-certificate/)

Data Science Courses
MSDS 400-DL Math For Data Scientists (1 Unit)
Students learn techniques for building and interpreting mathematical models of real-world phenomena in and across multiple disciplines,
including linear algebra, discrete mathematics, probability, and calculus, with an emphasis on applications in data science and data engineering. Provides students with a firm understanding or review of these fields of mathematics prior to enrolling in courses that assume understanding of mathematical concepts.

**Prerequisite:** None.

**MSDS 401-DL Applied Statistics with R (1 Unit)**
This course teaches fundamentals of statistical analysis. This includes evaluating statistical information, performing data analyses, and interpreting and communicating analytical results. Students will learn to use the R language for statistical analysis, data visualization, and report generation. Topics covered include descriptive statistics, central tendency, exploratory data analysis, probability theory, discrete and continuous distributions, statistical inference, correlation, multiple linear regression, contingency tables, and chi-square tests. Selected contemporary statistical concepts, such as bootstrapping, are introduced to supplement traditional statistical methods. Recommended prior course: MSDS 400-DL.

**Prerequisite:** None.

**MSDS 402-DL Introduction to Data Science (1 Unit)**
This course introduces the field of data science, which combines business strategy, information technology, and modeling methods. The course reviews the benefits and opportunities of data science, as well as organizational, implementation, and ethical issues. The course provides an overview of modeling methods, analytics software, and information systems. It discusses business problems and solutions for traditional and contemporary data management systems, and the selection of appropriate tools for data collection and analysis. The course also reviews approaches to business research, sampling, and survey design.

**Prerequisite:** None.

**MSDS 403-DL Data Science in Practice (1 Unit)**
This is a case study course that gives students an opportunity to gain experience solving business problems and applying core skills needed for data science technical and leadership roles. The course provides an introduction to digital transformation, industry use cases, designing and measuring analytics projects, data considerations, data governance, digital trust and ethics, enterprise architecture and technology platforms, and organizational change management. Students act as data scientists, as strategists and leaders, evaluating alternative analytics projects and solving digital transformation challenges. Students learn how to apply a step-by-step development process, creating digital transformation roadmaps and addressing real-world business problems.

**Prerequisite:** None.

**MSDS 410-DL Supervised Learning Methods (1 Unit)**
This course introduces traditional statistics and data modeling for supervised learning problems, as employed in observational and experimental research. With supervised learning there is a clear distinction between explanatory and response variables. The objective is to predict responses, whether they be quantitative as with multiple regression or categorical as with logistic regression and multinomial logit models. Students work on research and programming assignments, exploring data, identifying appropriate models, and validating models. They utilize techniques for observational and experimental research design, data visualization, variable transformation, model diagnostics, and model selection.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 411-DL Unsupervised Learning Methods (1 Unit)**
This course introduces data modeling for studies in which there is no clear distinction between explanatory and response variables. Students learn how to explain relationships among many continuous variables in terms of underlying dimensions, as with principal components and factor analysis. They find lower-dimensional representations for multivariate cross-classified data, as with log-linear models. They visualize data with traditional multidimensional scaling and t-distributed stochastic neighbor embedding. And they identify groups of variables and objects that are similar to one another, as with cluster analysis and biclustering. Students work on research and programming assignments, exploring multivariate data and methods.

**Prerequisite:** MSDS 400-DL and MSDS 401-DL.

**MSDS 413-DL Times Series Analysis and Forecasting (1 Unit)**
This course covers analytical methods for time series analysis and forecasting. Specific topics include the role of forecasting in organizations, exploratory data analysis, stationary and non-stationary time series, autocorrelation and partial autocorrelation functions, univariate autoregressive integrated moving average (ARIMA) models, seasonal models, Box-Jenkins methodology, regression models with ARIMA errors, multivariate time series analysis, and non-linear time series modeling including exponential smoothing methods, random forest analysis, and hidden Markov modeling. Recommended prior course: MSDS 410-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 420-DL Database Systems and Data Preparation (1 Unit)**
This course introduces data management and data preparation with a focus on applications in large-scale analytics projects. Students learn about relational databases, the relational model, the normalization process, and structured query language. They learn about data cleaning and integration, and database programming for extract, transform, and load operations. Students work with unstructured data, indexing and scoring documents for effective and relevant responses to user queries. They learn about graph data models and query processing. Students write programs for data preparation and extraction using various data sources and file formats. Recommended prior programming experience or MSDS 430-DL.

**Prerequisite:** MSDS 402-DL or MSDS 403-DL.

**MSDS 422-DL Practical Machine Learning (1 Unit)**
The course introduces machine learning with business applications. It provides a survey of statistical and machine learning algorithms and techniques including the machine learning framework, regression, classification, regularization and reduction, tree-based methods, unsupervised learning, and fully-connected, convolutional, and recurrent neural networks. Students implement machine learning models with open-source software for data science. They explore data and learn from data, finding underlying patterns useful for data reduction, feature analysis, prediction, and classification.

**Prerequisite:** MSDS 400-DL, MSDS 401-DL, and MSDS 402-DL or MSDS 403-DL.

**MSDS 430-DL Python for Data Analysis (1 Unit)**
This course introduces core features of the Python programming language, demonstrating fundamental concepts in computer science. It provides an in-depth discussion of data representation strategies, showing how data structures are implemented in Python and demonstrating tools for data science and software engineering. Working on data analysis problems, students employ various programming paradigms, including functional programming, object-oriented programming, and data stream processing. Special attention is paid to the standard Python library and packages for analytics and modeling.

**Prerequisite:** None.

**MSDS 432-DL Foundations for Data Engineering (1 Unit)**
This course provides an overview of the discipline of data engineering. It introduces software and systems for data science and software
development as required in the design of data-intensive applications. Students learn about algorithms, data structures, and technologies for storing and processing data. Students gain experience with open-source software, text editors, integrated development environments, and cloud systems. Students employ best practices in software development, utilizing tools for syntax checking, testing, debugging, and version control. The course also introduces formal models for algorithm and system performance. Prerequisite:

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 420-DL or CIS 417-DL.

**MSDS 434-DL Analytics Application Engineering (1 Unit)**

This course covers programming components essential to the development of analytics applications. The focus is analytics software engineering. Students learn to develop desktop and client-server solutions. They learn about web-based solutions employing a variety of front-end and back-end system components. The course introduces machine learning operations and engineering. Students use cloud systems to package and distribute containerized computer software. They develop software, working on open-source programming, database, and systems integration projects. They employ best practices in software development. Recommended prior courses: MSDS 432-DL.

**Prerequisite:** 1) MSDS 400-DL and 2)MSDS 420-DL or CIS 417-DL.

**MSDS 436-DL Analytics Systems Engineering (1 Unit)**

This course introduces design principles and best practices for implementing large-scale systems for data ingestion, processing, storage, and analytics. Students learn about cloud-based computer architecture and scalable systems for data science. They evaluate performance and resource utilization in batch, interactive, and streaming environments. Students review protocols for application programming interfaces. They compare data models, resource requirements, and performance of applications implemented with relational versus graph database systems. Recommended prior course: MSDS 432-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 440-DL Real-Time Interactive Processing and Analytics (1 Unit)**

This course introduces application engineering and analytics within an integrated environment and full-stack development process. Students implement client-side, web-based applications using a model-view-controller framework. They use server-side systems for responding to website requests and database queries. They prepare indices for efficient, relevant search across large document collections. They find information in databases and document collections, make service and product recommendations, and detect anomalies or security violations. This is a case study and project-based course with a strong programming component.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 442-DL Real-Time Stream Processing and Analytics (1 Unit)**

This course introduces application engineering and analytics within stream and event processing environments. Students learn how to work with various data feeds and sources, including electronic sensors, monitoring continuous processes, observing communication traffic and social interaction, and tracking goods through production and distribution. Students implement stream processing solutions, providing high throughput and low latency. They use relational and graph databases. They analyze event logs and business processes. This is a case study and project-based course with a strong programming component.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 450-DL Marketing Analytics (1 Unit)**

This course reviews applications of data science in marketing, the strategic marketing process, and the design of marketing surveys and experiments. Students explore methods for understanding consumer preferences, market segments, and competitive brands and products. Students address problems in new product design and pricing. They study the marketing mix, highlighting the effects of advertising and promotion. And they are introduced to algorithms and methods for digital marketing. Recommended prior courses: MSDS 410-DL and MSDS 411-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 451-DL Financial Machine Learning (1 Unit)**

This course introduces applications of machine learning techniques to finance. Financial data presents special challenges to standard machine learning techniques, engendering significant adaptations. Topics include a basic introduction to finance, nuances of financial features engineering, techniques to avoid various biases during model training, and example applications such as meta-labeling. Recommended prior course: MSDS 413-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 452-DL Web and Network Data Science (1 Unit)**

This course shows how to acquire and analyze information from the web and reviews web analytics and search performance metrics. It introduces the mathematics of network science, including random graph, small world, and preferential attachment models. Students compute network metrics, analyzing structure and connections in information and social networks. They study user interactions through electronic communications and social media. They work with graph algorithms and graph databases. This is a case study and project-based course with a strong programming component.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 453-DL Natural Language Processing (1 Unit)**

This course reviews natural language processing with a focus on recent developments in computational linguistics and machine learning. Students work with unstructured and semi-structured text from online sources, document collections, and databases. Students learn how to parse text into numeric vectors and to convert higher dimensional vectors into lower dimensional vectors for subsequent analysis and modeling. Applications include speech recognition, semantic processing, text classification, search, recommendation systems, sentiment analysis, and topic modeling. This is a project-based course with extensive programming assignments.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 454-DL Applied Probability and Simulation Modeling (1 Unit)**

This advanced modeling course begins by reviewing probability theory and models. Students learn principals of random number generation and Monte Carlo methods for classical and Bayesian statistics. They are introduced to applied probability models and stochastic processes, including Markov Chains, exploring applications in business and scientific research. Students work with open-source and proprietary systems, implementing discrete event and agent-based simulations. This is a case study and project-based course with an extensive programming component. Recommended prior course: MSDS 460-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 455-DL Data Visualization (1 Unit)**
This course begins with a review of human perception and cognition, drawing upon psychological studies of perceptual accuracy and preferences. The course reviews principles of graphic design, what makes for a good graph, and why some data visualizations effectively present information and others do not. It considers visualization as a component of systems for data science and presents examples of visualizing categorical, hierarchical, relational, temporal and spatial data. It reviews methods for static and interactive graphics and introduces tools for building web-browser-based presentations. This is a project-based course with programming assignments.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 456-DL Sports Performance Analytics (1 Unit)**

An introduction to sports performance measurement and analytics, this course reviews roles of athletes at each position in sports selected by the instructor. With a focus on the individual athlete, the course discusses the development and use of accurate assessments and variability due to factors such as body type, climate, and training regimen. The course reviews athletic performance measurements, including jumping ability, running speed, agility, and strength. Students work with player on-field and on-court performance measures. The course utilizes exploratory data analysis, predictive modeling, and presentation graphics, showing real-world implications for athletes, coaches, team managers, and the sports industry.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 457-DL Sports Management Analytics (1 Unit)**

This course provides a comprehensive review of financial, statistical, and mathematical models as they relate to sports team performance, administration, marketing, and business management. The course gives students an opportunity to work with data and models relating to sports team performance, tactics, and strategy. Students employ modeling methods in studying player and team valuation, sports media, ticket pricing, game-day events management, loyalty and sponsorship program development, and customer relationship management. The course makes extensive use of sports business case studies.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 458-DL Artificial Intelligence and Deep Learning (1 Unit)**

An introduction to artificial intelligence, this course illustrates probability-rule-based generative models as well as discriminative models for learning from data. It reviews applications of artificial intelligence and deep learning in vision and language processing. Students learn best practices for building deep learning models for classification and regression. The learn about feature engineering, autoencoders, and strategies of unsupervised and semi-supervised learning, as well as reinforcement learning. This is a project-based course with extensive programming assignments.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 459-DL Knowledge Engineering (1 Unit)**

This course reviews methods for developing knowledge-based systems, providing examples of their use in intelligent applications and conversational agents. It uses of relational, document, and graph databases for storing information about relationships among words, people, places, events, and things. Students learn about knowledge representation and automated reasoning. They draw on logic programming and machine learning to build end-to-end knowledge-based applications for information extraction and question answering. Recommended prior courses: MSDS 453-DL and MSDS 458-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 460-DL Decision Analytics (1 Unit)**

This course covers fundamental concepts, solution techniques, modeling approaches, and applications of decision analytics. It introduces commonly used methods of optimization, simulation and decision analysis techniques for prescriptive analytics in business. Students explore linear programming, network optimization, integer linear programming, goal programming, multiple objective optimization, nonlinear programming, metaheuristic algorithms, stochastic simulation, queuing modeling, decision analysis, and Markov decision processes. Students develop a contextual understanding of techniques useful for managerial decision support. They implement decision-analytic techniques using a state-of-the-art analytical modeling platform. This is a problem and project-based course.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 460-DL Decision Analytics (1 Unit)**

This course covers fundamental concepts, solution techniques, modeling approaches, and applications of decision analytics. It introduces commonly used methods of optimization, simulation and decision analysis techniques for prescriptive analytics in business. Students explore linear programming, network optimization, integer linear programming, goal programming, multiple objective optimization, nonlinear programming, metaheuristic algorithms, stochastic simulation, queuing modeling, decision analysis, and Markov decision processes. Students develop a contextual understanding of techniques useful for managerial decision support. They implement decision-analytic techniques using a state-of-the-art analytical modeling platform. This is a problem and project-based course.

**Prerequisite:** (1) MSDS 400-DL and (2) MSDS 401-DL.

**MSDS 462-DL Computer Vision (1 Unit)**

This course reviews deep learning methods for vision. Students work with raw image files, including digital representations of photographs, handwritten documents, x-rays, and sensor images. They process image data, converting pixels into numeric tensors for subsequent analysis and modeling. The course illustrates real-world applications for visual exploration, object recognition, image classification, facial recognition, remote sensing, navigation, and medical diagnostics. This is a project-based course with extensive programming assignments. Recommended prior course: MSDS 458-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 464-DL Intelligent Systems and Robotics (1 Unit)**

This course introduces reinforcement learning as an approach to intelligent systems. It reviews Markov decision processes, dynamic programming, temporal difference learning, Monte Carlo and deep reinforcement learning, eligibility traces, and function approximation. Students implement intelligent agents, solving sequential decision-making problems. They develop, debug, train, and visualize the results of programs. They see how to integrate learning and planning. This is a case study and project-based course with a substantial programming component. Recommended prior course: MSDS 458-DL.

**Prerequisite:** (1) MSDS 420-DL or CIS 417-DL and (2) MSDS 422-DL or CIS 435-DL.

**MSDS 470-DL Analytics Entrepreneurship (1 Unit)**

This course prepares students for establishing and running a data-sciences-oriented entrepreneurial organization. Students learn how to transform an idea into a running organization, identifying the right data, analytics tools, and resources, and aligning with unmet market demands. Students learn about growing a network of people and obtaining capital assets, creating innovative intellectual property, sharpening unique competitiveness, and making product development and marketing choices. Analytic product and consulting services opportunities are
MSDS 480-DL Business Leadership and Communications (1 Unit)
This course introduces concepts of leadership and organizational behavior. It builds on the premise that leadership is learned and discusses how to drive effective change in organizations at various stages. Students spend three weeks on analytics-specific project management, in which they design an analytics project plan using an agile approach. They learn how to incorporate the cross-industry standard processes for data analysis and modeling. They practice executing plans in simulated business settings. Working on case studies and theory-based assignments, students see how to address leadership challenges unique to analytics organizations. The course focuses on developing effective communication strategies and presentations that resonate across business and technical teams.
Prerequisite: None.

MSDS 480-DL Business Leadership and Communications (1 Unit)
This course introduces concepts of leadership and organizational behavior. It builds on the premise that leadership is learned and discusses how to drive effective change in organizations at various stages. Students spend three weeks on analytics-specific project management, in which they design an analytics project plan using an agile approach. They learn how to incorporate the cross-industry standard processes for data analysis and modeling. They practice executing plans in simulated business settings. Working on case studies and theory-based assignments, students see how to address leadership challenges unique to analytics organizations. The course focuses on developing effective communication strategies and presentations that resonate across business and technical teams.
Prerequisite: None.
MSDS 499-DL Independent Study (1 Unit)
Topics vary from term to term.
Prerequisite: Vary by topic.

MSDS 579-0 Practicum (1 Unit)

MSDS 590-DL Thesis Research (1 Unit)
This final project is meant to represent the culmination of students’ experience in the program and must demonstrate mastery of the curriculum and ability to conduct sustained independent research and analysis. The project may be applied or may be a traditional scholarly paper, in both cases a write-up following the paper’s program-specific guidelines is required. Students must submit a proposal and secure a first reader in order to register; for further details students are advised to review the student handbook and contact their academic adviser.