DATA SCIENCE (DATA)

DATA1000 DATA ANALYSIS
Data is all around us, and the ability to analyze it has become a necessary skill in many disciplines. This course is designed to introduce students to the fundamental principles of data analysis and provide them with the necessary tools to explore and draw insights from complex data sets. Students learn how to use statistical methods to summarize and visualize data, and how to use a spreadsheet to clean and transform data. Studentsalso learn how to use various data analysis techniques, including regression analysis, and classification, to identify patterns and make predictions.(4 credits)

DATA3010 DATA MINING
This course will primarily focus on techniques used for solving data mining problems. Topics include data preprocessing; mining frequent patterns; associations, and correlations; text mining; graph mining; clustering; and recommender systems. In addition, the course will look at ethical implications of data mining. ***Prerequisites:*** *COMP3125 (4 credits) fall*

DATA5000 APPLIED LARGE LANGUAGE MODELS
In this course, students explore the real-world applications of large language models (LLMs) in various domains. They learn the fundamentals of LLMs, including model architectures like GPT-3, BERT, and T5, and their applications in tasks such as text generation, semantic search, and chatbots. Through hands-on projects, students utilize pre-trained models and fine-tune them for specific tasks to solve real-world problems. (4 credits)

DATA5500 SENIOR DESIGN
Students will form small teams and participate in the design and implementation of data-insightful solutions to large-scale problems. Problems will be chosen in consultation with the instructor. Each team is required to apply techniques and tools throughout the data analysis lifecycle and present the resulting knowledge via written documentation and oral presentation. ***Prerequisite:*** *Senior Status (4 credits) summer*

DATA6000 APPLIED STATISTICS FOR RESEARCH
This course introduces graduate students in data science and engineering to a set of supervised and unsupervised tools for modeling and understanding complex datasets. Topics covered include statistical learning, linear regression, point and interval estimation and hypothesis testing, analysis of variance, classification, linear model selection and regularization, support vector machines, and unsupervised learning. (3 credits) fall

DATA6100 DATA VISUALIZATION
This course focuses on strengthening students' core data-analysis and data-visualization mindset and skills and prepares students to contribute to any organization's data-inspired decision-making. Students will leave this course with improved abilities to articulate project objectives; blend, structure, and analyze datasets; and share findings using data visualizations that engage audiences through story, interactivity, and design principles grounded in an understanding of human perception. (3 credits) fall

DATA6150 DATA SCIENCE FOUNDATIONS
This course provides an overview of the field of data science and examines a variety of tools for solving data-insightful problems. The presented methodologies include numerical computing, data processing, visualization, and making predictions via machine learning. Bias and ethics in data science and their potential consequences are also discussed. (3 credits) fall

DATA6200 DATA MANAGEMENT
This course introduces the necessary concepts for comprehensive data management strategy, such as data collection, storage, and retrieval. Topics covered include database management systems, introduction to database theory, and management of data in an organizational environment, as well as big data technologies. (3 credits) spring

DATA6250 MACHINE LEARNING FOR DATA SCIENCE
This course introduces supervised and unsupervised machine learning techniques as well as main approaches to dimensionality reduction using minimal theory and hands-on examples. The course utilizes popular, yet simple and efficient tools to learn from data and to create production-ready models. Topics start with fundamental classification algorithms, such as linear regression, and progress to deep neural networks. ***Prerequisite:*** *DATA6150 (3 credits) spring*

DATA6300 ADVANCED TOPICS IN LARGE LANGUAGE MODELS
This course dives deep into the state-of-art research and advancements in large language models (LLMs) and their impact on natural language processing (NLP) in recent years. Through a meticulous exploration of seminal and contemporary literature, students grasp the technical foundations, emerging capabilities, fine-tuning strategies, system design considerations, and the ethical and security facets of LLMs. The course culminates in a research project, granting students hands-on experience and an opportunity to contribute to this dynamic field. (3 credits) fall

DATA6710 APPLIED DEEP LEARNING
The objective of this course is to democratize the popular field of artificial intelligence (AI) by designing deep learning architectures via open-source, popular, yet easy-to-use frameworks. The course covers the fundamentals of deep learning and best practices to develop state-of-the-art AI models to solve vision and sequence processing problems, as well as generative and adversarial techniques. ***Prerequisites:*** *DATA6150 (3 credits) summer*

DATA6900 CAPSTONE I
This course is the first of a two-course capstone sequence in which students apply their cumulative learning from their program to address a real-world problem. In the sequence students develop a data-informed solution using their analytical, statistical, and/or visualization skills to address complex social problems and to demonstrate integrated knowledge. This first course in the sequence focuses on quantitative and qualitative research methods, including methodology and design, with its main deliverable being an approved project proposal. ***Prerequisites:*** *DATA6150 or MGMT6150 (3 credits) spring*

DATA6950 CAPSTONE II
This course is the second half of the Data Science and Business Analytics capstone sequence. Students complete the project proposed in Capstone I, setting forth a constructive argument supported by an analysis of existing literature and synthesis of data. Upon completing the project, the students are to communicate their findings effectively to an audience, in oral, visual, and/or in written format. ***Prerequisite:*** *DATA6900 (3 credits) summer*