Data Science

The Department of Data Science is the newest addition to the Ying Wu College of Computing. It was founded by well established, prominent researchers and educators with outstanding track records in Artificial Intelligence, Machine Learning, High Performance Data Analytics, Security/Privacy/Ethics in Data Science, Health Data Science, Green Data Science, and Data Visualization. The Department of Data Science was founded in 2021.

Data Science combines powerful methods from Computer Science, Statistics, Artificial Intelligence and Machine Learning into a unique new blend of techniques for deriving valuable insights from Big Data. Data Science is an ideal choice for students who are interested in applying data processing methods to ever larger and more varied real-world data sets, including image, video, natural language and speech data that go substantially beyond traditional text and table data to solve real-world problems. The Department of Data Science closely collaborates with the Department of Mathematical Sciences and the Department of Computer Science. Students also can get involved in state-of-the-art research projects at the NJIT Institute for Data Science, where top notch scientists work with users to develop data-driven technologies to innovate the way the world works and lives.

The department offers B.S., M.S., and Ph.D. degree programs in Data Science, Artificial Intelligence and specialized interdisciplinary programs.

B

Bader, David, Distinguished Professor

Bhavsar, Jaini, University Lecturer

D

Dasgupta, Aritra, Assistant Professor

Du, Mengnan, Assistant Professor

G

Geller, James, Professor, Chair

I

Islam, Akm, University Lecturer

K

Kaur, Ravneet, University Lecturer

L

Li, Daming, Senior University Lecturer

M

Monogioudis, Pantelis, Professor of Practice

N

Nemane, Nikita, University Lecturer

P

Pethkar, Kaustubh, University Lecturer

Phan, Hai, Associate Professor

R

Rangamani, Akshay, Assistant Professor

Renda, Michael, Professor of Practice

Roshan, Usman, Associate Professor

W

Wang, Lijing, Assistant Professor

Wu, Chase, Professor

X

Xu, Mengjia, Assistant Professor

Y

Yusuf, Fatima, University Lecturer

Z

Zhang, Shuai, Assistant Professor

Programs

Artificial Intelligence - M.S. (http://catalog.njit.edu/graduate/computing-sciences/data-science/artificial-intelligence-ms/)

Data Science - M.S.– Computational Track (http://catalog.njit.edu/graduate/computing-sciences/data-science/data-science-ms/)

Data Sciences - Ph.D. (http://catalog.njit.edu/graduate/computing-sciences/data-science/data-science-phd/)

- Artificial Intelligence (http://catalog.njit.edu/graduate/computing-sciences/data-science/artificial-intelligence-cert/)
- Big Data Essentials (http://catalog.njit.edu/graduate/computing-sciences/data-science/big-data-essentials-cert/)
- Data Mining (http://catalog.njit.edu/graduate/computing-sciences/data-science/data-mining-cert/)
- Data Visualization (http://catalog.njit.edu/graduate/computing-sciences/data-science/data-visualization-cert/)

DS 633. Advances in Representation Learning. 3 credits, 3 contact hours.
Prerequisites: DS 675. The current era of data science requires identifying structure in high-dimensional signals. Transforming data into “good” representations can make the difference in the success or failure of downstream machine learning algorithms for classification, regression, or data generation. We will explore unsupervised learning algorithms that deal with linear and nonlinear structures and are also able to learn representations from multiple data modalities. We will use applications in computer vision, natural language processing, and audio processing to motivate our theory and algorithms.

DS 636. Data Analytics with R Program. 3 credits, 3 contact hours.
Prerequisites: Entry-level courses in programming, probability and statistics (e.g MATH 333, CS 280, or equivalent courses with permission of the instructor). This course teaches data analytics with R programming. The student will learn and gain basic analytic skills via this high-level language. The course covers fundamental knowledge in R programming. Popular R packages for data science will be introduced as working examples. The course also includes case studies on data analytics projects. As a core course in data science, it provides skills that are highly desirable for both industry and academic employers.

DS 637. Python and Mathematics for Machine Learning. 3 credits, 3 contact hours.
This course aims to equip students with foundational knowledge and practical skills in Python programming and mathematics as they relate to machine learning. By combining theory and practice, students will gain a solid understanding of the basics of Python programming, data analysis, probability, statistics, linear algebra and optimization, and be prepared for more advanced studies in machine learning. This course could be a valuable first step for anyone interested in pursuing a career or further studies in machine learning.

DS 639. Elec. Medical Records: Med Terminologies and Comp. Imp.. 3 credits, 3 contact hours.
This course presents a graduate introduction to Medical Informatics for Computer Science students covering (1) the design, use and auditing of medical terminologies, such as the Unified Medical Language System (UMLS) and the Systematized Nomenclature of Medicine (SNOMED); and (2) principles of Electronic Medical Records (EMR), Electronic Health Records (EHR) and Personal Health Records (PHR), including issues of privacy and security.

DS 642. Applications of Parallel Computing. 3 credits, 3 contact hours.
Prerequisites: Proficiency in (non-parallel) programming in a high level procedural language. This course will teach students how to design, analyze, and implement, parallel programs for high performance computational science and engineering applications. The course focuses on advanced computer architectures, parallel algorithms, parallel languages, and performance-oriented computing. Students will develop knowledge and skills to efficiently solve challenging problems in science and engineering, where very fast computers are required either to perform complex simulations or to analyze enormous datasets.

DS 644. Introduction to Big Data. 3 credits, 3 contact hours.
Prerequisites: Permission of the instructor. This course provides an in-depth coverage of various topics in big data from data generation, storage, management, transfer, to analytics, with focus on the state-of-the-art technologies, tools, architectures, and systems that constitute big-data computing solutions in high-performance networks. Real-life big-data applications and workflows in various domains (particularly in the sciences) are introduced as use cases to illustrate the development, deployment, and execution of a wide spectrum of emerging big-data solutions.

DS 650. Data Visualization and Interpretation. 3 credits, 3 contact hours.
The course will focus on training students with the knowledge of data visualization theory, techniques, and tools. Students will learn why and how visualization can be applied in the human-centered data science pipeline and the different uses of visualization, such as in exploratory data analysis, and in the communication of data-driven insights. They will gain practical experience in interpreting, critiquing, and comparing visualization techniques by using real-world data sets and case studies. Students will also develop interactive visualization interfaces as part of the class project. They will gain a broad understanding of how visualization can enhance trust and interpretation of machine learning models. The students will read and learn about recent progress in the areas of information visualization, visual analytics, and human-data interaction.

DS 669. Reinforcement Learning. 3 credits, 3 contact hours.
Prerequisites: Linear algebra, basic probability, basic calculus, computer programming, or approval of instructor. Experience with machine learning, artificial intelligence, or deep learning (e.g., CS 675, CS 670, CS 677) is recommended. This course covers current topics, key concepts, and classic and modern algorithms in reinforcement learning and contains both theory and applications. The topics include but are not limited to, Markov Decision Processes, exploration and exploitation, planning, value-based learning, policy gradient, etc. Students will present recent papers in reinforcement learning, and work on written and programming assignments and do a reinforcement learning project. After completing this course, students will be able to start using reinforcement learning for real world problems that can be specified as Markov Decision Processes.

DS 675. Machine Learning. 3 credits, 3 contact hours.
Prerequisites: Basic probability, linear algebra, computer programming, and graduate or undergraduate senior standing, OR approval of instructor. This course is an introduction to machine learning and contains both theory and applications. Students will get exposure to a broad range of machine learning methods and hands on practice on real data. Topics include Bayesian classification, perceptron, neural networks, logistic regression, support vector machines, decision trees, random forests, boosting, dimensionality reduction, unsupervised learning, regression, and learning new feature spaces. There will be several programming assignments, one course project, one mid-term and one final exam.

DS 677. Deep Learning. 3 credits, 3 contact hours.
Prerequisites: DS 675 or CS 675 or approval of the instructor. This course covers current topics in data science. The topics include but are not limited to parallel programming on GPU and CPU multi-cores, deep learning, representation learning, optimization algorithms, and algorithms for big datasets. Students will present recent papers in data science, work on programming assignments, and do a machine learning/deep learning/data science project.

DS 680. Natural Language Processing. 3 credits, 3 contact hours.
Prerequisites: DS 675 or DS 677 and instructor's approval. This course aims to teach how to process one of the fundamental data sources—natural language—with the help of deep learning techniques. The target of this course is to familiarize students with state-of-the-art language models, wide variety of tasks performed with these models and the fusion of these in deep learning architectures. This course will help students read advanced research papers on complex NLP concepts and theories, while the class project will help them apply NLP techniques to different domains.

DS 681. Deep Learning for Computer Vision. 3 credits, 3 contact hours.
Prerequisites: DS 675. This course projects the vast field of statistical learning using differential deep neural architectures onto the computer vision application space. Beginning with the fundamentals of computer vision, the course offers an extensive coverage on essential topics such as object detection, semantic segmentation, using Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The course extends on such fundamentals and treats computer vision in multimodal and generative settings enabling applications such as image captioning, visual question answering, and scene generation, using state-of-the-art models like Neural Radiance Fields (NeRF) and diffusion models. Students at the end of the course are well-equipped to design and deploy vision systems capable of complex tasks, from tracking and identifying objects in video streams to generating interactive responses based on visual prompts.

DS 683. Graph Neural Networks. 3 credits, 3 contact hours.
Prerequisites: DS 675. Graphs provide a natural framework for representing complex relationships between various objects. Graph Neural Networks (GNNs) have gained significant importance in both academic research and industrial applications. This course introduces GNNs and explores foundational concepts, algorithms, and diverse applications. Students will learn the fundamentals of graph theory, and key models, e.g., Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), advanced graph diffusion models, and integrations of GNNs with sequential models for temporal graph modeling. The course will cover practical applications across fields like social networks, bioinformatics, and finance, focusing on hands-on implementation and problem-solving. By the end, students will be skilled in designing and applying GNN models to real-world datasets.

DS 685. Artificial Intelligence for Robotics. 3 credits, 3 contact hours.
Prerequisites: DS 675. This course introduces students to the foundational concepts of AI as they apply to the complex world of robotics, emphasizing how AI algorithms enable robots to interpret sensor data, navigate dynamic environments, and interact intelligently with both objects and humans. The course explores key AI methods, including computer vision, natural language processing, path planning and reinforcement learning that give robots the ability to analyze, reason, and respond to the world around them. Students will engage with hands-on projects with a focus on real-world applications in areas like autonomous vehicles and industrial automation. By the end of the course, students will have gained a deep understanding of how AI enables autonomous robots to perform complex tasks, being instructed with natural language making them capable of working alongside humans in various settings.

DS 687. Artificial Intelligence for Temporal Data. 3 credits, 3 contact hours.
Prerequisites: None. However, prior coursework such as MATH 447, DS 675, DS 677, or CS 670 can provide valuable background for understanding the topics in this course. Additionally, a solid foundation in Python programming is essential for success in this course. Time series data appears across domains such as finance, climate science, and healthcare. This course offers an in-depth exploration of machine learning techniques for time series analysis and predictive modeling. Students will address the unique challenges of temporal data, delving into statistical methods, classical machine learning, and advanced deep learning models, including RNNs, LSTMs, CNNs, GNNs, and Transformers. The course emphasizes theoretical foundations, algorithmic implementation, and research-driven applications, preparing students to tackle complex real-world problems and contribute to cutting-edge developments in the field.

DS 688. Advanced Federated Machine Learning. 3 credits, 3 contact hours.
Prerequisites: DS 675. The increasing availability of data has greatly boosted the power of machine learning (ML). However, modern data generation (e.g., from personal devices, within hospitals) fundamentally changes ML pipelines. Unlike traditional pipelines that use centralized datasets collected from the web to train ML models, these new data modes result in heterogeneous siloed data residing in the devices or organizations that generated it.To make use of this decentralized data, we focus on collaborative and federated ML, which enable secure and trustworthy learning across multiple parties and diverse data sources.

DS 698. Special Emerging Topics. 3 credits, 3 contact hours.
Prerequisites: Determined by the nature of the topic area. The course covers special emerging topics in DS that are not offered as regular courses.

DS 700B. Master's Project. 3 credits, 3 contact hours.
Restrictions: Approval of the project advisor is required for registration. Approval of the project advisor is required for registration. Experimental and/or theoretical investigation of a relevant topic in data science. A written report must be submitted to the project advisor. The student cannot register in DS 700B more than once and the incomplete (I) grade is not allowed. Master’s students registering for the first time in Master’s Project must take simultaneously the INTD 799 (Responsible Conduct of Research) course.

DS 701B. Master's Thesis. 3 credits, 3 contact hours.
Restrictions: Approval of the thesis advisor is required for registration. Approval of the thesis advisor is required for registration. Experimental and/or theoretical investigation of a relevant topic in data science that can lead to a quality publication. A written thesis must be defended and approved by a committee of at least three faculty members. The student is expected to defend the thesis upon accrual of six thesis credits. Additional registration in DS 701B, beyond six credits, is required every semester until successful thesis defense (six credits count toward degree requirements and time limits apply). Master’s students registering for the first time in Master’s Thesis must take simultaneously the INTD 799 (Responsible Conduct of Research) course.

DS 701C. Master's Thesis. 6 credits, 6 contact hours.
Restrictions: Approval of the thesis advisor is required for registration. Approval of the thesis advisor is required for registration. Experimental and/or theoretical investigation of a relevant topic in data science that can lead to a quality publication. A written thesis must be defended and approved by a committee of at least three faculty members. The student must continue registering for three thesis credits (DS 701B) each semester until successful thesis defense (six credits count toward degree requirements and time limits apply).

DS 725. Independent Study I. 3 credits, 3 contact hours.
Approval of the academic advisor is required for registration. Students working on their PhD dissertation cannot register for both DS 725 and DS 726 with the same faculty. This special course covers areas of study in which one or more students may be interested but there is not sufficiently broad interest to warrant a regular course offering. Students may not register for this course more than once.

DS 726. Independent Study II. 3 credits, 3 contact hours.
Approval of the academic advisor is required for registration. Students working on their PhD dissertation cannot register for both DS 725 and CS 726 with the same faculty. This special course covers areas of study in which one or more students may be interested but there is not sufficiently broad interest to warrant a regular course offering. Students may not register for this course more than once.

DS 732. Theoretical Foundation of Machine Learning. 3 credits, 3 contact hours.
Prerequisites: DS 675. This course provides an advanced exploration of machine learning, emphasizing its mathematical foundations and the interplay between statistical and computational aspects. This course is designed for graduate-level students seeking to deepen their understanding of machine learning. It supports those aiming to contribute to learning theory and algorithm development. The course combines conceptual insights with quantitative analysis to build a strong foundation in the field. Key topics include classical results in generalization and optimization theory, recent advances in deep learning, unsupervised learning, and large language models (LLMs). Students will acquire tools to analyze and prove performance guarantees for learning methods, fostering a strong theoretical foundation and practical expertise in machine learning.

DS 786. Selected Topics in Data Science. 3 credits, 3 contact hours.
Prerequisites: As determined by nature of topic area. Introduction to selected topics in data science.

DS 789. Trustworthy Artificial Intelligence. 3 credits, 3 contact hours.
Prerequisites: DS 675 or approval of instructor. As machine learning (ML) systems are increasingly being deployed in real-world applications, it is critical to ensure that these systems are behaving responsibly and are trustworthy. That will lead to wider adoption of ML in real-world applications in practice. This course will provide a deep understanding of state-of-the-art ML methods designed to make AI more trustworthy to unforeseen faults, adversarial manipulation, and to violations of ethical norms in privacy and fairness. Students will gain an understanding of and experience in using a set of methods and tools for deploying transparent, ethically sound, and robust machine learning solutions. The course is also an excellent opportunity to conduct research on the security/privacy/trustworthiness in ML and find research topics for Ph.D. and M.S. theses.

DS 790A. Doctoral Dissertation & Research. 1 credit, 1 contact hour.
Corequisites: DS 791. Approval of the dissertation advisor is required for registration. Experimental and/or theoretical investigation of a relevant topic in data science. For PhD students who have successfully defended their dissertation proposal. The student must register in DS 790A every semester until successful dissertation defense. A written dissertation must be defended and approved by a committee of at least five members.

DS 791. Graduate Seminar. 0 credits, 1 contact hour.
Corequisite (for doctoral students only): DS 790. A seminar in which faculty, students, and invited speakers will present summaries of advanced topics in data science. In the course, students and faculty will discuss research procedures, dissertation organization, and content. Students engaged in research will present their own problems and research progress for discussion and criticism.

DS 792B. Pre-Doctoral Research. 3 credits, 3 contact hours.
Corequisites: DS 791. Approval of the dissertation advisor is required for registration. Preliminary experimental and/or theoretical investigation of a relevant topic in data science. For students who have passed the qualifying examination but have not defended the dissertation proposal. Permission is needed of the academic advisor as well for students who have completed the required coursework but have not passed the qualifying examination.