Data Science

The Department of Data Science is the newest addition to the Ying Wu College of Computing. Founded in 2021, it offers a B.S. Degree in Data Science. This is a new degree program that responds to a strong demand from employers for trained Data Scientists. Data is revolutionizing most industries and B.S. graduates in Data Science command high starting salaries.

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 and students can take advantage of many computer science and mathematical sciences offerings. The Department of Data Science offers its own two-semester capstone projects that are executed with industrial sponsors. 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.

B

Bader, David, Distinguished Professor

D

Dasgupta, Aritra, Assistant Professor

Du, Mengnan, Assistant Professor

G

Gaikwad, Nikita, Lecturer

Geller, James, Professor

I

Islam, Akm, University Lecturer

L

Li, Daming, Senior University Lecturer

M

Monogioudis, Pantelis, Professor of Practice

P

Pethkar, Kaustubh Lecturer

Phan, Hai, Assistant Professor

R

Renda, Michael, Professor of Practice

Roshan, Usman, Associate Professor

W

Wang, Lijing, Assistant Professor

Wu, Chase, Professor

X

Xu, Mengjia

Y

Yusuf, Fatima, University Lecturer

Z

Zhang, Shuai, Assistant Professor

Programs

Data Science (Computing Option)- B.S. (http://catalog.njit.edu/undergraduate/computing-sciences/data-science/data-science-bs/)

- Artificial Intelligence Minor (for DS and CS Majors) (http://catalog.njit.edu/undergraduate/computing-sciences/data-science/artificial-intelligence-minor-cs-ds-majors/)
- Artificial Intelligence Minor (for non-DS and non-CS Majors) (http://catalog.njit.edu/undergraduate/computing-sciences/data-science/artificial-intelligence-minor-non-cs-ds-majors/)

DS 100. Basic Foundations of Data Science. 3 credits, 3 contact hours (3;0;0).
Prerequisites: No prerequisites are required to enroll. Data Science (DS) and artificial intelligence (AI) systems are increasingly being deployed in real-world applications across domains, significantly impacting our daily and social lives. It is critical to ensure that students have a good understanding of the new era of the DS/AI-human-centric world. That will lead to a broader adoption of DS/AI in real-world applications in practice. This course will provide an insightful understanding of recent data science (DS) and artificial intelligence (AI) developments. Students will learn basic program skills in Python and enable a venture through basic building blocks of data structures, data collection, processing, and generating. Using data visualization tools, students will learn to analyze real-world datasets across domain applications. Students will explore the ethics of AI abstractly and comprehensively, ranging from societal risks, regulations, and responsible technologies in AI. Hands-on labs are developed to align basic knowledge and practical skills in DS/AI. As a result, the course is designed to offer an appropriate entry point into the vibrant world of DS/AI for students with no background in computing so that students can diversify and strengthen their career paths by using DS/AI appropriately and optimally.

DS 110. Basic Foundations of Artificial Intelligence. 3 credits, 3 contact hours (3;0;0).
This course introduces essential and basic artificial intelligence (AI) tools in Python, which has four units: sensor stream, AI operating system, AI cognitive core, and AI symbolic compositional models. Sensor Stream: Gathers raw data from the environment through senses. AI Operating System (AIOS): Manages basic AI cognitive and AI physiological processes, integrating sensory input and regulating AI behavior. AI Cognitive Core: Engages in higher-level reasoning, decision-making, and learning. AI Symbolic Compositional Models: Facilitates symbolic tasks, such as language, mathematics, and creativity, by combining abstract representations. Students will learn basic programming skills in Python and solve problems mimicking the basic building blocks of cognition – modern AI mapping.

DS 240. General Introduction to Data Science. 3 credits, 3 contact hours (3;0;0).
Prerequisites: CS 100 or CS 101 or CS 103 or CS 104 or CS 106 or CS 113 or CS 115 or BME 210 or BNFO 135 with a grade C or better. Restrictions: This course is not for DS and CS majors, DS or CS students need to take DS 340 or CS 301 instead. This course provides a basic, yet comprehensive coverage of the fundamental principles and practical applications of data science and artificial intelligence (AI). This course, intended for all majors at NJIT, provides an introduction to Data Science with reduced coding. The course progresses to help students build a solid foundation for data processing, computing, and analysis. Topics include data manipulation, visualization, big data ecosystem, machine learning, deep learning, trustworthy AI, AI ethics, and cutting-edge advancements such as large language models and AI for sciences. Hands-on work involves Python with popular libraries including Pandas, NumPy, and PyTorch. This course is not for DS and CS majors.

DS 340. Fundamentals and Principles of Data Science. 3 credits, 3 contact hours (3;0;0).
Prerequisites: CS 114 and (MATH 333 or MATH 341) with a grade C or better. Fundamentals and principles of data science familiarize students with the theories and techniques for data representation, manipulation, analysis, visualization, and interpretation. Topics include introduction to data preparation and preprocessing, data mining, anomaly detection, machine learning, statistical learning, data analysis and visualization, large language models, ethics, and popular data science tools and systems. Hands-on work will include Python with Pandas coding.

DS 400. Scientific Foundation of Machine Learning. 3 credits, 3 contact hours (3;0;0).
Prerequisites: (CS 100 or DS 100) and (CS 375 or CS 370). This course provides an advanced exploration of machine learning, emphasizing its mathematical foundations and the interplay between statistical and computational aspects. Designed for students seeking to deepen their understanding or advance the theory and development of learning algorithms, the course primarily focuses on conceptual insights at the undergraduate level. Key topics include classical results in statistical and computational learning 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 410. Federated Machine Learning and Applications. 3 credits, 3 contact hours (3;0;0).
Prerequisites: CS 375. 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 enables secure and trustworthy learning across multiple parties and diverse data sources.

DS 450. Data Visualization. 3 credits, 3 contact hours (3;0;0).
Prerequisites: CS 301 or DS 340, all with a grade C or better. The course provides students an introduction to computer graphics and the knowledge for designing, developing, and applying techniques for both information and volume visualization. Software tools such as Tableau and programing languages such as Python will be used to represent and interpret information in various visual forms, and volumetric visualization algorithms such as marching cubes and ray casting will be used for big data visualization of 3D datasets in scientific domains. Students will gain knowledge about theoretical design principles and apply them directly on real-world data, as part of assignments and course projects.

DS 480. Fundamentals and Applications of Graph Neural Networks. 3 credits, 3 contact hours (3;0;0).
Prerequisites: (CS 100 or DS 100) and CS 375. 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, biological networks, brain networks, and finance, focusing on hands-on implementation and problem-solving. By the end of the semester, students will be skilled in designing and applying GNN models to real-world datasets.

DS 485. Selected Topics in DS. 3 credits, 3 contact hours (3;0;0).
Restrictions: Junior standing and/or department approval. The study of new and/or advanced topics in an area of data science not regularly covered in any other DS course. The precise topics to be covered in the course, along with prerequisites, will be announced in the semester prior to the offering of the course. A student may register for no more than two semesters of Special Topics.

DS 487. Artificial Intelligence for Temporal Data. 3 credits, 3 contact hours (3;0;0).
Prerequisites: MATH 447, CS 375. Time series data appears across domains such as finance, climate science, healthcare, and many more. This course offers an in-depth exploration of artificial intelligence techniques for temporal data analysis and predictive modeling. Students will address the unique challenges of temporal data, delving into classical machine learning, and advanced deep learning models, including RNNs, LSTMs, CNNs, GNNs, and Transformers. The course emphasizes algorithmic implementation and research-driven applications, preparing students to tackle complex real-world problems and contribute to cutting-edge developments in the field.

DS 488. Independent Study in Data Science. 3 credits, 3 contact hours (3;0;0).
Restrictions: Open only to Data Science majors who have the prior approval of the department and the DS faculty member who will guide the independent study. Independent studies, investigations, research, and reports on advanced topics in data science. Students must prepare, in collaboration with their faculty mentor and in the semester prior to enrolling in this course, a detailed plan of topics and expected accomplishments for their independent study. This must have the approval of both the department and the faculty mentor. A student may register for no more than one semester of Independent Study.

DS 492. Data Science Capstone I. 3 credits, 3 contact hours (3;0;0).
Restrictions: Senior standing. The Data Science (DS) Capstone Project spans two semesters and is intended to provide a real-world project-based learning experience for seniors in the BS DS program. The overall objectives of this course are to investigate the nature and techniques of a data-oriented computing development project. Projects are provided by faculty members or industry partners, or proposed by students who wish to become entrepreneurs. In DS Capstone I, teams of project participants will carry out market research, identify appropriate data science problems, collect and preprocess the needed data, define performance metrics, perform risk analysis, and finish an overall design of their solution that integrates various data analytics techniques. The course instructor will mentor and evaluate all projects in conjunction with an entrepreneurship board of industry, faculty, and alumni advisors.

DS 493. Data Science Capstone II. 3 credits, 3 contact hours (3;0;0).
Prerequisites: DS 492 with a grade C or better. The Data Science (DS) Capstone Project spans two semesters and is intended to provide a real-world project-based learning experience for seniors in the BS DS program. The overall objectives of this course are to investigate the nature and techniques of a data-oriented computing development project. Projects are provided by faculty members or industry partners, or proposed by students who wish to become entrepreneurs. In DS Capstone II, teams of project participants will refine their design, implement and integrate component techniques into a complete software solution, present data analysis results, evaluate the system performance, and validate the proposed solution. The course instructor will mentor and evaluate all projects in conjunction with an entrepreneurship board of industry, faculty, and alumni advisors.