

# Ph.D. in Data Science

---

## I. Admission Requirements

Prospective applicants are expected to have software development experience, computational skills, and an understanding of statistical methods. The minimum requirements for admission to the PhD program are within the guidelines and policies approved by the University and include:

- A Bachelor's degree in data science, computer science, informatics, mathematics/statistics, engineering, or another closely related discipline (as approved by the Ph.D. director) from a college or university accredited in the United States, or its equivalent, with a minimum overall GPA of 3.5 out of 4.0.
- GRE scores are required.
- International student applicants shall demonstrate proficiency in English if English is not their first language, following the NJIT admission standard. Exemptions can be granted to applicants who have earned (or will earn, before enrolling at NJIT) a U.S. bachelor's, master's, or doctoral degree from a university of recognized standing in a country in which all instruction is provided in English.
- Prepared students shall have a good background in programming and data structures (e.g. NJIT CS 280 and CS 435), advanced Calculus (e.g. NJIT Math 211), and Probability and Statistics (e.g. Math 333/341). Admitted students lacking competencies in one or more of these areas shall consult with the academic advisor to take relevant preparatory courses. Students might be required to enroll in a relevant Certificate Program at NJIT and will only be admitted with a GPA of 3.0 or higher in the Certificate Program.

Applicants to the program are expected to be drawn from diverse backgrounds, including both domestic and international students, and students from diverse ethnic groups, as well as under-represented groups. The program aspires to achieve a fair and equitable balance in gender and ethnicity representation that reflects the population of the United States and overcomes the skewed ratios of the past, in full agreement with federal and state laws.

## II. Degree Requirements

### II.1 Course Requirements

The courses include core courses, elective courses, and courses for conducting research. All core courses are listed in Table DR-1. The definition of "core courses" in this document is that they are offered by the Department of Data Science or the Department of Mathematical Sciences and are considered especially relevant to Data Science and are recommended to students as such. Table DR-2 provides a partial list of the elective courses available to program students. In addition to the listed elective courses, a student may take other special topic courses, at most two of which can be counted as electives, subject to the approval of the academic advisor.

Course descriptions for the core courses and elective courses listed in Tables DR-1 and DR-2 can be found in the University Graduate Catalog online at <http://catalog.njit.edu/graduate/>. Courses listed are offered by the Ying Wu College of Computing (YWCC), the College of Science and Liberal Arts (CSLA), and the Newark College of Engineering. These colleges are happy to collaborate with DS/MATH, providing regular course offerings and accommodating the Ph.D. students in the Data Science program.

Courses for conducting research include: DS 790A - Doctoral Dissertation & Research; DS 791 - Doctoral Seminar; DS 792B - Pre-Doctoral Research, described below:

#### **DS 792B - Pre-Doctoral Research**

Ph.D. students who pass the Qualifying Exam must then register for 3 credits of pre-doctoral research per semester until they successfully defend the dissertation proposal.

#### **DS 791 - Doctoral Seminar**

Ph.D. students are required to register each semester for a zero-credit Graduate Seminar. Attendance and participation in the Seminar are required of all students.

#### **DS 790A - Doctoral Dissertation & Research**

PhD students who successfully defend the dissertation proposal must then register for the one-credit dissertation course each semester until they complete all the degree requirements.

Students may take courses simultaneously with the DS 791 or DS 792B course as per academic advisor and dissertation advisor/committee recommendation.

Statistics Track:

- Ph.D. students with a recognized Baccalaureate degree are required to take ten (NJIT minimum: eight) 600-level or 700-level 3-credit courses (30 credits) of coursework beyond the Baccalaureate degree as well as four additional 700-level 3-credit courses (12 credits), for a total of fourteen (NJIT minimum: twelve) 3-credit courses (42 credits).
- Ph.D. students with a recognized Master's degree or equivalent are required to take seven graduate courses, of which four should be 700-level 3-credit courses (21 credits). (NJIT minimum: four 700-level courses.)
- Master's project (course MATH 700), Master's thesis (course MATH 701), or more than two independent study courses (courses MATH 725 and MATH 726) cannot be used to satisfy these coursework requirements.
- Students will be required to take DS 675 (Machine Learning), MATH 644 (Regression), and MATH 631 (Linear Algebra).
- All required courses can be substituted by courses of equal difficulty, if the Ph.D. advisor and the Ph.D. directors in *both* tracks agree to them in writing. For example, if a student has already taken an equivalent course to a required course, then a substitute will be determined.

#### Computing Track:

- Ph.D. students with a recognized Baccalaureate degree are required to take ten (NJIT minimum: eight) 600-level or 700-level 3-credit courses (30 credits) of coursework beyond the Baccalaureate degree as well as four additional 700-level 3-credit courses (12 credits), for a total of fourteen (NJIT minimum: twelve) 3-credit courses (42 credits).
- Ph.D. students with a recognized Master's degree or equivalent are required to take seven graduate courses, of which four should be 700-level 3-credit courses (21 credits). (NJIT minimum: four 700-level courses.)
- Master's project (course DS 700), Master's thesis (course DS 701), or more than two independent study courses (courses DS 725 and DS 726) cannot be used to satisfy these coursework requirements.
- Students will be required to take DS 675 (Machine Learning), MATH 644 (Regression), and DS 644 (Introduction to Big Data)
- All required courses can be substituted by courses of equal difficulty, if the Ph.D. advisor and the Ph.D. directors in *both* tracks agree to them in writing. For example, if a student has already taken an equivalent course to a required course, then a substitute will be determined.

In addition, both tracks require a two-part qualifying exam, a proposal document, a proposal defense, a dissertation document, and dissertation defense. Publication requirements will be defined by the track directors. Students may request to transfer between the tracks prior to passing their qualifying exams.

With approval by the academic advisor and dissertation advisor, a student is allowed to take elective courses based on the dissertation topic. The following are examples of potential dissertation areas and possible elective courses appropriate for the student's program of study. *Note: please see course listing in Table DR-2 or visit the online Graduate Catalog as cited above for further course information and descriptions.*

- A student with interest in Machine Learning or related areas may choose elective courses such as CS 732 Advanced Machine Learning, etc. Potential research topics may include, but are not limited to, algorithm development for clustering, dimensionality reduction, reinforcement learning, and machine learning in Natural Language Processing.
- A student with interest in Statistics may choose MATH 787 Non-parametric statistics, MATH 786 Large Sample Theory and Inference. Potential research topics may include, but are not limited to, machine learning, uncertainty quantification, statistical learning, and data mining.
- A student with interest in Data Visualization may choose DS 650 Data Visualization. Potential research topics may include, but are not limited to, visualization techniques for explainable AI, visual analytics for human-machine trust, and communicative visualization design.
- A student with interest in High Performance Computing may choose DS 642 Applications of Parallel Computing, CS 668 Parallel Algorithms, CS 750 High Performance Computing, etc. Potential research topics may include, but are not limited to, Real-world algorithms, Numerical computing, Scalable Systems, High Performance Data Analytics, Modeling & Simulation

#### II.2 Other Requirements

Students are expected to have their research findings published in high quality peer-reviewed academic conference proceedings and journals at a volume that is considered the established standard in their subfield of Data Science.

Students are also required to attend and participate in Data Science research seminars every semester and are encouraged to attend other research seminars across campus. Seminar attendance will be monitored and recorded by the Ph.D. program academic advisor. Students in the Computing Track should attend research seminars in YWCC. Students in the Statistics Track should attend research seminars in the Department of Mathematical Sciences. Students in both concentrations should attend all Data Science related seminars.

To continue in the PhD program, a student must fulfill the following requirements/milestones. Failure to satisfy these requirements may result in probation or dismissal from the program:

- Maintain a cumulative GPA of 3.0 or better. Students will need a cumulative GPA of 3.5 if they wish to be considered for financial support of any kind.
- End of year one: Student must take the written part of the PhD qualifying exam. Upon the approval of the PhD program director, student must file a program of study that lists the courses to be taken and the timeline of study. Policy for repetition in case of failure will be publicized at the time of admission into the Ph.D. program.

- Students are recommended to choose a dissertation advisor as soon as possible, but no later than 3 months after passing the qualifying exam.
- End of the third semester: Student must present a review of current research literature in the chosen area of research in a written format and in an oral defense in front of a committee of three faculty members.
- Any change to the program of study must be approved by the Ph.D. program director and the dissertation advisor (if chosen).
- End of year two: student must have passed the qualifying exam.
- End of year three: student must have a dissertation committee established and the dissertation proposal must be successfully defended.
- Student must attend at least 70% of the research seminars for six semesters.
- The dissertation should be presented in writing and should be orally defended by the end of the fourth or fifth year, and must be defended at latest by the end of the sixth year in the PhD program. Students who cannot defend their dissertation by the end of the sixth year will be dismissed from the program.

Students are responsible for prerequisites of elective courses.

Code	Title	Credits
<b>Core Courses</b>		
DS 675:	Machine Learning	3
DS 644:	Introduction to Big Data	3
DS 636:	Data Analytics with R Programming	3
DS 677:	Deep Learning	3
DS 642:	Applications of Parallel Computing	3
DS 632:	Advanced Data Management	3
DS 650:	Data Visualization	3
DS 680:	Natural Language Processing	3
DS 725:	Independent Study in Data Science I	3
DS 726:	Independent Study in Data Science II	3
DS 790A:	Doctoral Dissertation & Research	3
DS 791:	Graduate Seminar	0
DS 792:	Pre-Doctoral Research	3
DS 786:	Special topics seminar in Data Science	3
MATH 644	Regression Analysis Methods	3
MATH 660	Introduction to statistical Computing with SAS and R	3
MATH 691	Stochastic Processes with Applications	3
MATH 611	Numerical Methods for Computation	3
MATH 678	Stat Methods in Data Science	3
MATH 699	Design and Analysis of Experiments	3
MATH 665	Statistical Inference	3
MATH 662	Probability Distributions	3
MATH 631	Linear Algebra	3
<b>Elective Courses (require approval)</b>		
CS 630	Operating System Design	
CS 631	Data Management System Design	
CS 634	Data Mining	
CS 656	Internet and Higher-Layer Protocols	
CS 670	Artificial Intelligence	
CS 610	Data Structures and Algorithms	3
CS 732	Advanced Machine Learning	3
CS 750	High Performance Computing	3
CS 645	Security and Privacy in Computer Systems	3
CS 602	Java Programming	3
CS 608	Cryptography and Security	3
CS 643	Cloud Computing	3
CS 647	Counter Hacking Techniques	3

CS 648	Cyber Sec Investigations & Law	3
CS 696	Network Management and Security	3
CS 708	Advanced Data Security and Privacy	3
ECE 601	Linear Systems	3
ECE 673	Random Signal Analysis	3
IE 650	Advanced Topics in Operations Research	3
IE 687	Healthcare Enterprise Systems	3
IE 688	Healthcare Sys Perfor Modeling	3
IS 634	Information Retrieval	3
IS 665	Data Analytics for Info System	3
IS 682	Forensic Auditing for Computing Security	3
IS 684	Business Process Innovation	3
IS 687	Transaction Mining and Fraud Detection	3
IS 688	Web Mining	3
MATH 787	Non-Parametric Statistics	3
MATH 786	Large Sample Theory and Inference	3
MATH 768	Probability Theory	3
MATH 763	Generalized Linear Models	3
MATH 707	Advanced Applied Mathematics IV: Special Topics	3
MATH 717	Inverse Problems and Global Optimization	3
MATH 761	Statistical Reliability Theory and Applications	3
MATH 659	Survival Analysis	3
MATH 680	Advanced Statistical Learning	3
MATH 683	High Dimensional Stat Inferenc	3
PHYS 621	Classical Electrodynamics	3
PHYS 641	Statistical Mechanics	3
PHYS 611	Adv Classical Mechanics	3
CHEM 658	Advanced Physical Chemistry	3
CHEM 714	Pharmaceutical Analysis	3
ME 625	Introduction to Robotics	3
ME 616	Matrix Methods in Mechanical Engineering	3
CE 611	Project Planning and Control	3
PTC 628	Analyzing Social Networks	3