DATA SCIENCE

Developed through a collaborative effort between the Departments of Statistics (http://statistics.columbian.gwu.edu/), Mathematics (http://math.columbian.gwu.edu/), Physics (http://physics.columbian.gwu.edu/), Economics (http://economics.columbian.gwu.edu/), Geography and Environment (http://geography.columbian.gwu.edu/), and Political Science (http://politicalscience.columbian.gwu.edu/), the Data Science Program offers the dual master of science in data science and graduate certificate in data science. The program teaches students to understand data and contribute important insights with the goal of changing the way in which people live, work, and communicate. Through a structured curriculum that provides foundational knowledge as well as application skills, students learn how to confront the most complex problems facing government and private industry

Visit the Data Science Program website (https://datasci.columbian.gwu.edu/) for additional information.

UNDERGRADUATE

Bachelor's program

- Bachelor of Science with a major in data science (https://bulletin.gwu.edu/arts-sciences/data-science/bs/)

Combined program

- Dual Bachelor of Science in an approved Columbian College program and Master of Science in the field of data science (https://bulletin.gwu.edu/arts-sciences/data-science/bs-ms/)

Minor

- Minor in data science (https://bulletin.gwu.edu/arts-sciences/data-science/minor/)

GRADUATE

Master's program

- Master of Science in the field of data science (https://bulletin.gwu.edu/arts-sciences/data-science/ms/)

Combined program

- Dual Master of Science in the field of data science and graduate certificate in geographic information systems (https://bulletin.gwu.edu/arts-sciences/data-science/ms-cert-gis/)

CERTIFICATE

Graduate certificate program

- Data science (https://bulletin.gwu.edu/arts-sciences/data-science/certificate/)

FACULTY

Professor: S. Kundu (*Program Director*)

Associate Professor: A. Jafari, Y. Huang

Assistant Professors: A. Walker, H. Shatnawi, J. Yin, S. Burnett, S. Majhi

COURSES

Explanation of Course Numbers

- Courses in the 1000s are primarily introductory undergraduate courses
- Those in the 2000s to 4000s are upper-level undergraduate courses that can also be taken for graduate credit with permission and additional work assigned
- Those in the 6000s and 8000s are for master's, doctoral, and professional-level students
- The 6000s are open to advanced undergraduate students with approval of the instructor and the dean or advising office

DATS 1001. Data Science for All. 3 Credits.
Cross-disciplinary perspective on topics ranging from an introduction and overview of standard data science tools through to frontier research topics in real-world systems. R is introduced.

DATS 2101. Ethical Life in a Digital World. 3 Credits.
Introduction to ethical perspectives and policies in an increasingly digital world. Topics include generation, recording, curation, processing, sharing, and use of data; algorithms; programming; hacking; and professional codes. Credit cannot be earned for this course and AMST 3625, ANTH 3625.

DATS 2102. Data Visualization for Data Science. 3 Credits.
Working with data and code to engage data science questions. Students develop coding, data visualization, and data presentation skills. Prerequisites: DATS 1001; STAT 1051 or STAT 1053 or STAT 1111 or STAT 1127 and CSCI 1012 or ISTM 3119 or STAT 1129. Recommended background: Students are required to have knowledge of Python.

DATS 2103. Data Mining for Data Science. 3 Credits.
Basic concepts, principles, methods, implementation techniques, and applications of data mining, with a focus on Python and data mining algorithms. Overview of data mining techniques and skills to explore and analyze data. Prerequisites: CSCI 1012, DATS 1001, MATH 1232, and (STAT 1051 or STAT 1053 or STAT 1111 or STAT 1127).

DATS 2104. Data Warehousing for Data Science. 3 Credits.
Fundamental concepts of databases and data warehousing. Database management; extract, transform, load (ETL) processes; and SQL. Ethical considerations, including privacy, data stewardship, and database security Prerequisites: CSCI 1012 and DATS 1001.

DATS 4001. Data Science Capstone. 3 Credits.
Capstone experience for data science majors. Application of theoretical knowledge and practical skills gained in major courses to a real-world problem. Review of ethical issues and current topics in data science. Restricted to data science majors. Prerequisites: DATS 1001, DATS 2101, DATS 2102, DATS 2103, and DATS 2104.

DATS 6001. Algorithm Design for Data Science. 3 Credits.
Theory and implementation of the most important problems in algorithm design. Tailored to the needs of non-computer science students.

DATS 6101. Introduction to Data Science. 3 Credits.
Basic techniques of data science. Algorithms for data mining, basics of statistical modeling, and concepts, abstractions, and practical techniques.

DATS 6102. Data Warehousing. 3 Credits.
Fundamentals and practical applications of data warehousing, including planning requirements, infrastructure, design, and maintenance. Prerequisites: STAT 2118 or permission of the instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics.

DATS 6103. Introduction to Data Mining. 3 Credits.
Concepts, principles, and techniques related to data mining; strengths and limitations of various data mining techniques, including classification, association analysis, and cluster analysis.

DATS 6202. Machine Learning I: Algorithm Analysis. 3 Credits.
Practical approach to fundamentals of algorithm design associated with machine learning; techniques of statistical and probability theory, combinatorial optimization, and factor graph and graph ensemble as used in machine learning. Prerequisites: DATS 6101 and DATS 6103. Credit cannot be earned for this course and PHYS 6620.

DATS 6303. Deep Learning and Fundamentals of AI. 3 Credits.
Focus on the implementation of state-of-the-art deep learning techniques on GPUs, first through presenting and implementing deep network architectures, then by describing how networks can be trained and analyzed within frameworks. Prerequisites: DATS 6101. Recommended background: Prior completion of any one of MATH 2233 or equivalent; time series modeling and analysis; machine learning; or linear algebra and stochastic system.

DATS 6304. Geographic Information Systems I. 3 Credits.
Fundamentals of cartography; geographic data structure and geographic information systems. Same As: GEOG 6304.

DATS 6311. Bayesian Methods in Data Science. 3 Credits.
Introduction to Bayesian data analysis. Parameter estimation (using formal analysis, grid approximation, and Markov chain Monte Carlo), hierarchical models, generalized linear models, JAGS, and Stan. Prerequisites: DATS 6101 and DATS 6103.

DATS 6312. Generative AI and Natural Language Processing. 3 Credits.
Foundational knowledge and cutting-edge skills in language analysis and data science, progressing systematically from classical techniques to advanced AI applications. Prerequisites: DATS 6202 and DATS 6103.

DATS 6313. Time Series Analysis and Modeling. 3 Credits.
Understanding, analyzing, modeling, and predicting time-series datasets. Fundamental concepts covered include stochastic systems, estimation theory, time series analysis and model validation. Based in Python programming. Restricted to graduate students in the data science program. Prerequisites: DATS 6101 and DATS 6103. Recommended background: science, mathematics, or statistics.

DATS 6401. Visualization of Complex Data. 3 Credits.
This course is a practical approach to fundamentals of data visualization specifically for data science professional. It covers all significant topics, including graphics, discrete and continuous variables, clustering and classification. Restricted to candidates for the MS or graduate certificate in data science; permission of the instructor may be substituted. Prerequisites: DATS 6101, DATS 6102, and DATS 6103.

DATS 6450. Topics in Data Science. 3 Credits.
Topics vary by semester. May be repeated for credit provided topic differs. See department for more details. Restricted to students in the master's and graduate certificate in data science programs. Restricted to students in the master's and graduate certificate programs in data science. Prerequisites: DATS 6101 or permission of the instructor.

DATS 6499. Data Science Applied Research. 1-3 Credits.
Students conduct research projects under the supervision of the instructor. Project topics build on the knowledge and skills acquired during the data science program. Permission of the instructor required prior to enrollment.

DATS 6501. Data Science Capstone. 3 Credits.
Practical application of the knowledge and skills acquired during the master's program. Capstone team projects are chosen in consultation with the instructor. In addition to the specific prerequisite courses, completion of five pre-approved data science courses is required prior to enrollment. Restricted to students in their final semester of the MS in data science program. Prerequisites: DATS 6101, DATS 6102 and DATS 6103.

DATS 6810. Applied Statistics and Data Analysis in Physics. 3 Credits.
Statistical inference applied to physical science data; modern statistical methods; informative visualizations of data and inferred trends. Analytical and practical skills for physical data analysis and interpretation using solid statistical methods. Prerequisites: MATH 2184, MATH 2233, PHYS 2023.