Data Science, B.S.

Degree Offered
• Bachelor of Science

Nature of the Program
Data science is an interdisciplinary field with roots in applied mathematics, statistics and computer science. The Bachelor of Science in Data Science seeks to meet the increased employment demands across many industries and research fields.

Data Science majors will develop quantitative and computational skills to solve real-world problems. For example, data scientists are responsible for creating and maintaining dashboards in a pandemic, predicting traffic patterns to improve driver safety and helping apps like Uber Eats optimize food delivery. Students can customize the degree to fit their interests by selecting a focus area of their choice to create a degree with practical applications.

Working with their academic advisers, data science students will take classes in a discipline related to their interests and career goals. Students can choose their area of emphasis among a variety of areas including the social sciences, humanities, and sciences. Examples, include astronomy, biology, criminology, geography, geology, GIS, physics, public health, psychology, and sociology.

FACULTY

DIRECTOR OF THE SCHOOL OF MATHEMATICAL AND DATA SCIENCES
• Earl Scime - Ph.D. University of Wisconsin, Madison
  Areas: fusion energy, space plasma physics, industrial plasma physics, plasma diagnostics, neurosciences imaging, magnetic reconnection, robotics, STEM education

PROFESSOR
• Snehalata Huzurbazar - Ph.D. (Colorado State University, Fort Collins)
  Areas: Statistics, Data Sciences

ASSISTANT PROFESSOR
• Srinjoy Das - Ph.D. (University of California San Diego)
  Areas: Data Sciences

Admissions
• First Time Freshmen are admitted to the major directly. For the timely completion of the degree, it is recommended that students have a minimum MATH ACT of 22, a MATH SAT of 540, or an ALEKS score of 45. Students must have a placement into a Math course to receive a complete schedule for their first semester. Students without placement may not be competitive to remain in the major
• Students transferring from another WVU major with fewer than 29 credits must have completed MATH 126 with a grade of C- or higher; students who have completed 30 or more credits must have completed MATH 154 or MATH 155 with C- or higher and have earned a 2.0 overall GPA.
• Students transferring from another institution with fewer than 29 credits must have completed MATH 126 with a grade of C- or higher; students who have completed 30 or more credits must have completed MATH 154 or MATH 155 with C- or higher and have earned a 2.0 overall GPA.

ADMISSIONS REQUIREMENTS FOR 2023-2024
The Admissions Requirements above will be the same for the 2023-2024 Academic Year.

Major Code: 14E7

General Education Foundations
Please use this link to view a list of courses that meet each GEF requirement. (http://registrar.wvu.edu/gef/)

NOTE: Some major requirements will fulfill specific GEF requirements. Please see the curriculum requirements listed below for details on which GEFs you will need to select.

General Education Foundations
F1 - Composition & Rhetoric
ENGL 101 Introduction to Composition and Rhetoric
& ENGL 102 and Composition, Rhetoric, and Research
or ENGL 103 Accelerated Academic Writing
Data Science, B.S.

Departmental Requirements for the B.S. in Data Science

• Capstone Requirement: The university requires the successful completion of a Capstone course. Data Science majors must complete DSCI 480.

• Writing and Communication Skills Requirements: Data Science Bachelor of Science students fulfill the Writing and Communication Skills requirement by completing ENGL 101 and ENGL 102 (or ENGL 103), and two additional SpeakWrite Certified Courses™:

• Calculation of the GPA in the Major: A minimum GPA of 2.5 across all classes applied to the Data Science Major Requirements is required. If a class is repeated, the second attempt will be included in the calculation of the GPA, unless it is subject to the D/F repeat policy.

• Advanced Coursework: As part of the major requirements, and in connection with their advisor, students will complete additional upper division coursework in a concentration of their choosing. Nine of the twelve credit hours must be at the 300-level or above.

• Benchmark Expectations: For details, for the Data Science Degree Progress tab.

Curriculum Requirements

University Requirements

ECAS B.S. Requirements

Departmental Requirements

Data Science Major Requirements

Total Hours

University Requirements

General Education Foundations (GEF) 1, 2, 3, 4, 5, 6, 7, and 8 (31-37 Credits)

Outstanding GEF Requirements 1, 4, 5, 6, and 7

DSCI 191 First-Year Seminar

General Electives

Total Hours

ECAS Bachelor of Science Requirements

Global Studies & Diversity Requirement

Mathematics Requirements

Fulfilled by major requirement

Science Requirement

Please see the Eberly College of Arts and Sciences’ Bachelor of Science (B.S.) tab.

Departmental Requirements

STEM FOUNDATIONS:

Basic Core Requirements: 19

MATH 153 Calculus 1a with Precalculus
MATH 154 and Calculus 1b with Precalculus
<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 155</td>
<td>Calculus 1</td>
</tr>
<tr>
<td>MATH 156</td>
<td>Calculus 2</td>
</tr>
<tr>
<td>CS 110</td>
<td>Introduction to Computer Science</td>
</tr>
<tr>
<td>&amp; CS 111</td>
<td>and Introduction to Data Structures</td>
</tr>
<tr>
<td>STAT 215</td>
<td>Introduction to Probability and Statistics</td>
</tr>
</tbody>
</table>

**Data Science Foundational Science Requirement**

Select one pair:

1. **Biological Sciences**
   - BIOL 115 & 115L Principles of Biology and Principles of Biology Laboratory
   - BIOL 117 & 117L Introductory Physiology and Introductory Physiology Laboratory

2. **Chemical Sciences**
   - CHEM 115 & 115L Fundamentals of Chemistry 1 and Fundamentals of Chemistry 1 Laboratory
   - CHEM 116 & 116L Fundamentals of Chemistry 2 and Fundamentals of Chemistry 2 Laboratory

3. **Physical Sciences**
   - PHYS 101 & 101L Introductory Physics 1 and Introductory Physics 1 Laboratory
   - PHYS 102 & 102L Introductory Physics 2 and Introductory Physics 2 Laboratory

4. **Chemical Sciences**
   - CHEM 117 & 117L Principles of Chemistry 1 and Principles of Chemistry 1 - Laboratory
   - CHEM 118 & 118L Principles of Chemistry 2 and Principles of Chemistry 2 - Laboratory

5. **Physical Sciences**
   - PHYS 111 & 111L General Physics 1 and General Physics 1 Laboratory
   - PHYS 112 & 112L General Physics 2 and General Physics 2 Laboratory

**Total Hours** 27

**Data Science Major Requirements**

**Mathematics Core**

- MATH 251 Multivariable Calculus
- MATH 303 Introduction to the Concepts of Mathematics
- MATH 378 Discrete Mathematics
- MATH 441 Applied Linear Algebra

**Statistics Core**

- STAT 312 Intermediate Statistical Methods
- STAT 445 Data Analysis

**Computer Science Core**

- CS 320 Analysis of Algorithms
- DSCI 301 Databases for Data Science

**Data Science Core**

- DSCI 101 Introduction to Data Science
- DSCI 221 Reproducible Data Science using R
- DSCI 222 Data Science Workflows using Python
- DSCI 310 Statistical Machine Learning 1
- DSCI 311 Statistical Machine Learning 2
- DSCI 410 Big Data in Practice: Cloud and Parallel Computing
- DSCI 450 Current Topics in Data Science

**Total Hours** 27
Data Science Advanced Electives

In consultation with an advisor, students will complete a concentration in a discipline of their choice such as Sociology, Geography, Biology or others. Students are welcome to propose concentrations that draw on their interests from the humanities, social sciences, or STEM fields where big data are collected and analyzed to provide new insights.

Capstone

<table>
<thead>
<tr>
<th>Course</th>
<th>Hours</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI 480</td>
<td>3</td>
<td>Capstone in Data Science</td>
</tr>
</tbody>
</table>

Total Hours 62

Suggested Plan of Study

First Year

<table>
<thead>
<tr>
<th>Fall</th>
<th>Hours</th>
<th>Spring</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI 191</td>
<td>1</td>
<td>DSCI 221</td>
<td>4</td>
</tr>
<tr>
<td>DSCI 101</td>
<td>3</td>
<td>CS 111 (B.S. First Area 2)</td>
<td>4</td>
</tr>
<tr>
<td>CS 110 (B.S. First Area 1)</td>
<td>4</td>
<td>MATH 156 (B.S. Second Area 1 Course 1; F8)</td>
<td>4</td>
</tr>
<tr>
<td>MATH 155 (F3)</td>
<td>4</td>
<td>F5</td>
<td>3</td>
</tr>
<tr>
<td>F4</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15 15

Second Year

<table>
<thead>
<tr>
<th>Fall</th>
<th>Hours</th>
<th>Spring</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI 222</td>
<td>3</td>
<td>DSCI 301</td>
<td>3</td>
</tr>
<tr>
<td>STAT 215 (F8 course 2)</td>
<td>3</td>
<td>MATH 441</td>
<td>3</td>
</tr>
<tr>
<td>MATH 303</td>
<td>3</td>
<td>STAT 312</td>
<td>3</td>
</tr>
<tr>
<td>MATH 251 (B.S. Second Area 2)</td>
<td>4</td>
<td>GEF 6</td>
<td>3</td>
</tr>
<tr>
<td>DSCI Foundational Science Elective (B.S. Third Area 1; F2)</td>
<td>4</td>
<td>DSCI Foundational Science Elective 1 (B.S. Third Area 2; F8 course 3)</td>
<td>4</td>
</tr>
</tbody>
</table>

17 16

Third Year

<table>
<thead>
<tr>
<th>Fall</th>
<th>Hours</th>
<th>Spring</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI 310</td>
<td>3</td>
<td>DSCI 311</td>
<td>3</td>
</tr>
<tr>
<td>STAT 445</td>
<td>3</td>
<td>MATH 378</td>
<td>3</td>
</tr>
<tr>
<td>CS 320</td>
<td>3</td>
<td>ECAS Global Studies and Diversity Requirement (F 7)</td>
<td>3</td>
</tr>
<tr>
<td>ENGL 101 (GEF 1)</td>
<td>3</td>
<td>ENGL 102 (GEF 1)</td>
<td>3</td>
</tr>
<tr>
<td>DSCI Advanced Science Elective 1</td>
<td>3</td>
<td>DSCI Advanced Science Elective 2</td>
<td>3</td>
</tr>
</tbody>
</table>

15 15

Fourth Year

<table>
<thead>
<tr>
<th>Fall</th>
<th>Hours</th>
<th>Spring</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSCI 410</td>
<td>3</td>
<td>DSCI 480</td>
<td>3</td>
</tr>
<tr>
<td>DSCI 450</td>
<td>3</td>
<td>Advanced Data Science Elective 4</td>
<td>3</td>
</tr>
<tr>
<td>DSCI Advanced Science Elective 3</td>
<td>3</td>
<td>General Elective</td>
<td>3</td>
</tr>
<tr>
<td>General Elective</td>
<td>3</td>
<td>General Elective</td>
<td>3</td>
</tr>
<tr>
<td>General Elective</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15 12

Total credit hours: 120

Degree Progress

- By the beginning of a student's third regular semester (fall or spring), they should have completed either MATH 154 or MATH 155 with a C- or better.
- During the first four regular semesters (fall and spring) in the major, student must complete their foundational mathematics courses through MATH 441, CS 110 and CS 111, and DSCI 101, DSCI 221, and DSCI 222.
- A minimum cumulative and major GPA of a 2.0 must be maintained. Students who do not meet this benchmark will be removed from the major.
### Major Learning Outcomes

**DATA SCIENCE**

**Learning Outcome 1:** Students will communicate data science workflows in both written and oral forms.

**Outcome 1.1** Students will demonstrate their ability to develop and use appropriate data science techniques to address ‘science’ (subject matter) topics and questions.

**Outcome 1.2** Students will communicate the biases and other implications of the data and analysis.

**Outcome 1.3** Students will prepare a clear and concise written project and orally present a data science workflow and analysis effectively and professionally.

**Learning Outcome 2:** Students will understand and demonstrate the programming and technological aspects of a data science workflow

**Outcome 2.1** Students will develop workflows using the languages and platforms common in data science practice (e.g., R and Python, Rstudio and JupyterLab)

**Outcome 2.2** Students will demonstrate their ability to acquire and manipulate data via a variety of platforms (e.g., databases to cloud computing)

**Outcome 2.3** Students will demonstrate their ability to use technologies for collaboration (e.g., Git and GitHub)

**Learning Outcome 3:** Students will demonstrate their ability to visualize and model data

**Outcome 3.1** Students will demonstrate visualization of data from simple plots for smaller data sets to visualizations for big data

**Outcome 3.2** Students will demonstrate their ability to use current machine learning and other data science modeling methods appropriately and understand the underlying statistical and mathematical concepts.

---

**DSCI 101. Introduction to Data Science. 3 Hours.**

Introduction and overview of this interdisciplinary field and the skills needed to work as a data scientist. Provides students basic experience in acquiring data, performing very simple analyses, and gaining an elementary understanding of data science.

**DSCI 191. First-Year Seminar. 1-3 Hours.**

Engages students in active learning strategies that enable effective transition to college life at WVU. Students will explore school, college and university programs, policies and services relevant to academic success. Provides active learning activities that enable effective transition to the academic environment. Students examine school, college and university programs, policies and services.

**DSCI 209. Data Science Pipelines with Python and R. 3 Hours.**

PR: DSCI 101 with a minimum grade of C- and MATH 124 or higher (up to MATH 156) with a minimum grade of C-. Development of workflow or computer programs to import, clean, transform, model and visualize data. Using data from different disciplines, students will program in Python and R as they develop these data science pipelines and present their results.

**DSCI 221. Reproducible Data Science using R. 4 Hours.**

PR: DSCI 101 and CS 110 with a minimum grade of C- in each. Introduction to programming in R and to using RStudio, and using the tidyverse set of packages to learn the basics of a data science pipeline needed to import, clean, transform, visualize and model large amounts of data.

**DSCI 222. Data Science Workflows using Python. 3 Hours.**

PR: DSCI 221 with a minimum grade of C-. Continuation of DSCI 221. Introduction to programming in Python, to the basics of building a data science pipeline. Students develop projects using data from various sources to develop and refine their Python skills. Also teaches the basics of terminal mode and use of bash.

**DSCI 301. Databases for Data Science. 3 Hours.**

PR: DSCI 209 with a minimum grade of C-. Focuses on understanding relational or categorical data structures associated with databases in a data science pipeline and acquiring data from existing databases using R and Python.

**DSCI 309. Applied Machine Learning. 3 Hours.**

PR: DSCI 209 with a minimum grade of C-. Statistical machine learning methods for supervised and unsupervised learning will be introduced via applications. Specifically, linear regression, methods for classification, resampling, model choice, dimension reduction and clustering will be covered with a conceptual understanding and their implementation using R and Python.

**DSCI 310. Statistical Machine Learning 1. 3 Hours.**

PR: DSCI 222 and STAT 312 and MATH 441 with a minimum grade of C- in each. Focuses on a conceptual understanding of the methods and their implementation using R and Python. Covers linear regression; classification methods (logistic regression, linear discriminant analysis and K-nearest neighbors); resampling methods (cross-validation and bootstrap); model choice methods (subset and stepwise selection, shrinkage methods); dimension reduction methods (principal components analysis).
DSCI 311. Statistical Machine Learning 2. 3 Hours.
PR: DSCI 310 with a minimum grade of C-. Continuation of DSCI 310. Covers statistical machine learning methods that are not strictly linear, such as models based on splines, tree-structures, support vector machines and unsupervised methods. Emphasizes a conceptual understanding and application of the methods using R and Python.

DSCI 409. Advanced Case Studies in Data Science. 3 Hours.
PR: DSCI 309 with a minimum grade of C-. The course covers advanced methods through case studies. Four main topics will be computing and analyzing data using the high performance computing, and case studies with natural language processing, real-time streaming data and imaging data. Concurrently students will acquire data from their own major and put together a data science pipeline and analysis for their final project.

DSCI 410. Big Data in Practice: Cloud and Parallel Computing. 3 Hours.
PR: DSCI 311 with a minimum grade of C-. Extends the R “tidyverse” data manipulation and machine learning pipelines to relational database tables; big data; network data; streaming data. Students will develop their abilities from using RStudio locally on a laptop to using it on a server, with technologies such as Spark.

DSCI 450. Current Topics in Data Science. 3 Hours.
PR: DSCI 311 with a minimum grade of C-. Exploration of timely current topics where data science is used; exploration and discussion of biases and other aspects of decisions made as a result of data science tools.

DSCI 480. Capstone in Data Science. 3 Hours.
PR: DSCI 410 and DSCI 450 with a minimum grade of C- in each. Integration and application of the skills and methods acquired through the program to a real data set through group project (development of a data science pipeline).