



# Data Science Program

## Program of Study

The ten course units for the Data Science Program are divided into three categories:

### 1. Foundations (two course units)

- **Probability:** ENM 503 Intro to Probability & Statistics or STAT 510 Probability or MATH 546 Advanced Probability
- **Programing Languages & Techniques (PL):** CIT 590 Programming Languages & Techniques or CIT 591 Introduction to Software Development

*If students have taken these courses as part of another program, this may be waived. For submatriculants, the probability requirement can be waived with successful completion of ESE 301 or STAT 430; the programming requirement can be waived with successful completion of CIS 120.*

*In lieu of these courses, students may take Technical Electives, and are encouraged (but not required) to take a course from Bucket C in lieu of Probability, and a course from Bucket B in lieu of PL.*

### 2. Core Requirements (three course units)

- **Mathematical Foundations:** STAT 512 Mathematical Statistics or CIS 515 Linear Algebra/Optimization or Computational Learning Theory CIS 625
- **Big Data Analytics:** CIS 545 Big Data Analytics
- **Mining and Learning:** CIS 519 Intro to Machine Learning or CIS 520 Machine Learning or STAT 571 Modern Data Mining

### 3. Technical & Depth Area Electives (five course units)

Students must choose courses from 3 *different buckets*, one bucket of which can be a 2 semester sequence of thesis/practicum. Two of the courses must represent a depth sequence, which could be the thesis/practicum or (for bucket options B-H) two courses, one of which builds on the other (e.g. is a prerequisite).

#### **BUCKETS for Technical & Depth Area Electives**

##### **A. Thesis/Practicum (two course units)**

Register for two credits of CIS 597/Master's Thesis or two credits of CIS 599/Master's Independent Study. Suggestions for projects will be posted on a website, or students should

feel free to find their own. Students will be mentored jointly by the Program Director and by an advisor in the area of the project, and must receive approval by Faculty Director.

#### **B. Data-Centric Programming**

- Software Systems (CIS 505)
- Software Engineering (CIS573)
- Computer Systems Programming (CIT595)
- Advanced Programming (CIS 552)
- Internet and Web Systems (CIS 555)
- Programming and Problem Solving (CIS559)

#### **C. Statistics, Mathematical Foundations**

- Numerical Methods (ENM502)
- Linear Algebra/Optimization (CIS515)
- Complex Analysis (AMCS510)
- Introduction to Optimization Theory (ESE504)
- Regression Analysis (STAT621)
- Stochastic Processes (STAT533)
- Bayesian Methods (STAT 542 )
- Convex Optimization (ESE605)
- Information Theory (ESE674)

#### **D. Data Collection, Representation, Management and Retrieval**

- Databases (CIS550)
- Sample Survey Methods (STAT920)
- Observational Studies (STAT921)

#### **E. Data Analysis, Artificial Intelligence**

**Note:** *Students should take at most two of CIS519/520, STAT571 and ESE545.*

- Computational Linguistics (CIS530)
- Computer Vision (CIS580, CIS581)
- Advanced Topics in Computer Vision (CIS680)
- Computational Learning Theory (CIS625)
- Data Mining: Learning from Massive Datasets (ESE545)
- Modern Data Mining (STAT571)
- Advanced Topics in ML (CIS700)
- Forecasting and Time-Series Analysis (STAT910)
- Algorithms (CIS502, CIS677, CIT596)
- AI (CIS521)
- Learning in Robotics (ESE650)
- Modern Regression for the Social, Behavioral and Biological Science (STAT 974)

#### **F. Modeling**

- Simulation Modeling and Analysis (ESE603)
- Control of Systems (ESE505)
- Data-driven Modeling and Probabilistic Scientific Computing (ENM540)

## **G. Biomedicine**

- Brain-Computer Interfaces (BE 521)
- Network Neuroscience (BE 566)
- Modeling Biological Systems (BE 567)
- Bioinformatics (STAT 953)
- Computational Neuroscience (PHYS 615)

## **H. Social/Network Science**

- Econometrics (ECON705, 706, 721, 722)
- MKTG476/776 Applied Probability Models in Marketing

