UCONN
UNIVERSITY OF CONNECTICUT

Office of the Provost
Anne D'Alleva, Ph.D.
Provost and Executive Vice President
for Academic Affairs

February 22, 2023

TO: Members of the Board of Trustees

FROM: Anne D'Alleva, Ph.D.
Provost and Executive Vice President for Academic Affairs

RE: Bachelor of Science in Data Science & Engineering

RECOMMENDATION:

That the Board of Trustees approve a new Bachelor of Science in Data Science and Engineering in the School of Engineering.

BACKGROUND:

In the past few years, demand for data scientists and engineers has skyrocketed. The latest job report from the career website dice.com lists "Data Scientists" and "Data Engineers" as 5th and 6th occupations by job posting volume, with 44.9% and 42.2% year-over-year growth in the first ten months of 2022, respectively. The latest projections from the US Bureau of Labor Statistics suggest that the strong demand for data scientists will continue into the next decade, with 40,500 new "Data Scientist" positions (35.8% growth) anticipated between 2021 and 2031.

UConn's College of Liberal Arts and Sciences (CLAS) recently developed a B.S. in Statistical Data Science and a B.A. degree in Applied Data Analysis. The School of Engineering proposes to complement the CLAS degrees with a new 120 credits B.S. in Data Science and Engineering. While the CLAS degrees emphasize training in statistical aspects of Data Science and application of Data Science methods to specific social science domains, the proposed curriculum for the Data Science and Engineering degree will provide graduates with strong computing competencies for designing and building systems for collecting, storing, and analyzing data at scale. The new major will be offered by the CSE Department at the main campus in Storrs, but students will be able to complete the first year of the major at branch campuses before transferring to Storrs.

The curriculum includes 33 credits of required Data Science and Engineering courses. A newly developed course on "Introduction to Data Science and Engineering" will introduce students to the principles of data acquisition, management, integration, predictive modeling, and visualization. Required courses will provide rigorous training in computer programming and fundamental Data Science and Engineering topics including algorithms, data structures, databases, statistical inference, predictive modeling, big data analytics, and machine learning. They will also cover data security, privacy, and ethics. All Data Science and Engineering majors will also be required to complete a year-long senior design project where they apply their Data Science and Engineering skills to design and implement solutions to real-world data-intensive problems. It's expected that once fully established, this program would graduate approximately 100 students per year.

352 MANSFIELD ROAD, UNIT 1086
STORRS, CT 06269-1086
PHONE: 860.486.4037
EMAIL: anne.dalleva@uconn.edu
WEB: provost.uconn.edu

UCONN

Proposal for a

Bachelor of Science in Data Science and Engineering

School of Engineering
University of Connecticut

1. Background and Rationale

Triggered by the challenge of dealing with an exponentially growing volume of data, the field of data analysis envisioned by Tukey 60 years ago [*] has dramatically evolved over the past decade into an interdisciplinary field commonly referred to as Data Science. In the past few years demand for data scientists and engineers has skyrocketed. The latest job report from the career website dice.com (Table 1) lists "Data Scientists" and "Data Engineers" as 5th and 6th occupations by job posting volume, with 44.9% and 42.2% year-over-year growth in the first ten months of 2022, respectively. The latest projections from the US Bureau of Labor Statistics[†] suggest that the strong demand for data scientists will continue into the next decade, with 40,500 new "Data Scientist" positions (35.8% growth) anticipated between 2021 and 2031.

Table 1. Year-to-Year growth for the top 10 tech occupations by job posting volume in the November 2022 Dice tech job report.

Rank	Occupation	YoY Change
1	Software Engineers	+28.4%
2	Business Analysts	+21.0%
3	Systems Engineers	+31.4%
4	Data Analysts	+47.7%
5	Data Scientists	+44.9%
6	Data Engineers	+42.2%
7	Software Developers	+1.0%
8	Electrical Engineers	+48.8%
9	DevOps Engineers	+9.7%
10	Java Developers	-19.4%

To meet this demand many US universities including UConn's peer and aspirant institutions have developed undergraduate degrees in Data Science and/or Data Science and Engineering. The College of Liberal Arts and Sciences at UConn has recently developed a BS degree in Statistical Data Science and a BA degree in Applied Data Analysis. We propose to complement the CLAS degrees with a new 120 credits BS degree in Data Science and Engineering. While the CLAS degrees emphasize training in statistical aspects of Data Science and application of Data Science methods to specific social science domains, the proposed curriculum for the Data Science and Engineering degree will provide graduates with strong computing competencies for designing and building systems for collecting, storing, and analyzing data at scale. The new major will be offered by the CSE Department at the main campus in Storrs, but students will be able to complete the first year of the major at branch campuses before transferring to Storrs.

---

[*] John W. Tukey, The Future of Data Analysis, Ann. Math. Statist. 33(1): 1-67 (March, 1962). DOI: 10.1214/aoms/1177704711

[†] https://www.bls.gov/emp/tables/fastest-growing-occupations.htm

The curriculum for the proposed Data Science and Engineering major, detailed in next section, has been informed by the report on "Computing Competencies for Undergraduate Data Science Curricula" released in 2021 by the Data Science Task Force convened by the Association for Computing Machinery's Education Board[‡]. The curriculum has been designed carefully to meet the proposed program criteria for Data Science, Data Analytics and similarly named computing programs[§] developed by ABET's Computing Accreditation Commission (CAC) and CSAB, the lead society for computing accreditation. The proposed ABET criteria are expected to become effective following the ABET Computing Area Delegation Meeting in the fall of 2022.

Under Criterion 3 (Student Outcomes) the ABET proposal requires that graduates of the program have an ability to:

1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
3. Communicate effectively in a variety of professional contexts.
4. Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
5. Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.
6. *Apply theory, techniques, and tools throughout the data science lifecycle and employ the resulting knowledge to satisfy stakeholders' needs. [DS]*

The first five student outcomes are general requirements for all computing-related programs, while the last one is specific to Data Science, Data Analytics and similarly named computing programs.

Under Criterion 5 (Curriculum), in addition to the general criteria curriculum requirements,[**] the ABET proposal includes the following:

At least 45 semester credit hours (or equivalent) of data science course work that must cover:

1. *Fundamental data science lifecycle topics:*
    a) *Data acquisition and representativeness*
    b) *Data management*
    c) *Data preparation and integration*
    d) *Data analysis*
    e) *Model development and deployment*
    f) *Visualization and communication of the knowledge obtained from the data*
2. *Concepts that span and are applied to the data science lifecycle:*
    a) *Data ethics including legitimate use and algorithmic fairness*
    b) *Governance including privacy, security, and stewardship*

[‡] https://www.acm.org/binaries/content/assets/education/curricula-recommendations/dstf_ccdsc2021.pdf
[§] https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2022-2023/

[**] Mathematics appropriate to the discipline and at least 30 semester credit hours or equivalent of up-to-date coverage of fundamental and advanced computing topics that provide both breadth and depth, including (1) Techniques, skills, and tools necessary for computing practice; (2) Principles and practices for secure computing; and (3) Local and global impacts of computing solutions on individuals, organizations, and society.

    c) *Statistical and mathematical topics including inference, modeling, linear algebra, probability and optimization*
    d) *Computing, including data structures, and algorithms*
3. *Advanced data science coursework that provides depth.*
4. *Coverage of at least one application area to provide a context for data science activities.*
5. *A major project that incorporates an application area and requires integration and application of knowledge and skills acquired in earlier course work.*

The math and science requirements for the new degree are comparable but more specific than the current math and science requirements for the BS in Computer Science. In particular, Multivariate Calculus is required for Data Science and Engineering majors (while Computer Science majors can take Differential Equations as an alternative) since it provides recommended preparation for the required course in Machine Learning. Similarly, Statistical Methods - which covers basic probability distributions, point and interval estimation, tests of hypotheses, correlation and regression, analysis of variance, experimental design, and non-parametric procedures - is specifically required for Data Science and Engineering majors whereas Computer Science majors can take several probability and mathematical statistics courses as alternatives.

The curriculum includes 33 credits of required Data Science and Engineering courses. These courses will provide students with the knowledge and skills required throughout the data science lifecycle. A newly developed course on "Introduction to Data Science and Engineering" will introduce students to the principles of data acquisition, management, integration, predictive modeling, and visualization. Required courses will provide rigorous training in computer programming and fundamental Data Science and Engineering topics including algorithms, data structures, databases, statistical inference, predictive modeling, big data analytics, and machine learning. They will also cover data security, privacy, and ethics. All Data Science and Engineering majors will also be required to complete a year-long senior design project where they apply their Data Science and Engineering skills to design and implement solutions to real-world data-intensive problems.

The proposed curriculum also includes 12 credits of Data Science and Engineering electives, for a total of 45 credits of Data Science and Engineering courses. Elective courses provide students the opportunity to acquire advanced Data Science and Engineering competencies such as artificial intelligence and data mining, software engineering, numerical computing, and stochastic modeling. Students will also be able to become familiar with various data-intensive scientific and engineering domains such as bioinformatics, biomedical engineering, chemical and material engineering, and other emerging areas. The list of elective courses is expected to expand over time with newly developed courses in Data Science and Engineering as well as new application domains.

Next section details the curriculum for the new degree. Catalog descriptions of the required Data Science and Engineering courses as well as a course schedule template are included in the Appendix.

2. Curriculum

2.1 Math and Science Requirements

- MATH 2110Q (Multivariable Calculus) and MATH 2210Q (Applied Linear Algebra)
- STAT 3025Q (Statistical Methods)
- One two-semester laboratory course sequence from either chemistry or physics: Chemistry: CHEM 1127Q–1128Q, 1137Q–1138Q, or 1147Q–1148Q; Physics: PHYS 1401Q–1402Q, 1501Q–1502Q, or 1601Q–1602Q
- One additional science course from the following list (but not in the same department as the two semester sequence): BIOL 1107, 1108, or 1110; CHEM 1127Q, or 1128Q; PHYS 1401Q, 1402Q, 1502Q, 1601Q, or 1602Q; ERTH 1050, or ERTH 1051 and 1052

2.2 Required Data Science and Engineering Courses

Course number	Credits	Course Name	Notes
CSE 1010	3	Introduction to Computing for Engineers	
CSE 2050	3	Data Structures and Object-Oriented Design	
CSE 2500	3	Introduction to Discrete Systems	
CSE 2600	3	Introduction to Data Science and Engineering	New course
CSE 3000	1	Contemporary Issues in Computer Science and Engineering	
CSE 3140	2	Cybersecurity Lab	
CSE 3500	3	Algorithms and Complexity	
CSE 4502	3	Big Data Analytics	
CSE 4701	3	Principles of Databases	
CSE 4820	3	Introduction to Machine Learning	
CSE 4939W	3	Senior Design I	
CSE 4940	3	Senior Design II	
Total	33		

2.3 Elective Data Science & Engineering Courses

A minimum of four (4) courses totaling 12 credits from the following list:

- CSE 2102 (Introduction to Software Engineering)
- CSE 3400 or CSE 5850 (Introduction to Cyber-Security)
- BME 3401 (Introduction to Computational and Systems Biology)
- CSE 3800 or CSE 5800 (Bioinformatics)
- CSE 3802 or ECE 3431 (Numerical Methods in Scientific Computation)
- STAT 3965 or MATH 3170 (Elementary Stochastic Processes)
- ECE 4131 (Introduction to Digital Signal Processing)
- ECE 4132 (Image Processing Systems Laboratory)
- CSE 4705 (Artificial Intelligence)
- BME 4810 (Machine Learning Methods for Biomedical Signal Analysis)
- CSE 5520 (Data Visualization and Communication)
- CSE 5713 (Data Mining)
- CSE 5820 (Machine Learning)

2.4 Free Electives

Additional elective courses to reach a minimum of 120 credits.

3. Enrollment Projections and Required Resources

Given the increasing interest for Data Scientist and Data Engineer careers (Figure 1) coupled with the high demand from employers (Table 1) we expect that a substantial number of incoming students will be interested in pursuing a BS in Data Science and Engineering. We anticipate that once the new program is fully established it will graduate around 100 Data Science and Engineering majors per year. This is supported by the fact that the "Computational Data Analytics" concentration for the Computer Science and Computer Science and Engineering majors is currently the second most popular concentration, attracting over 20% of the majors that declare a specialized concentration. The addition of the new major in Data Science and Engineering is expected to attract some of the students that would have otherwise pursued a CS or CSE degree with a "Computational Data Analytics" concentration, and if necessary, this concentration may need to be eliminated.

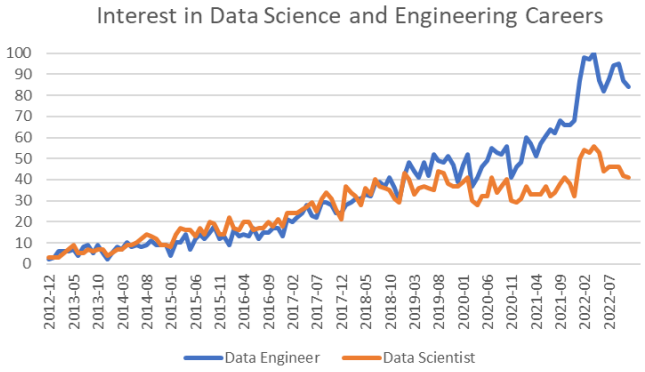

Figure 1. Google trends search term popularity under the Jobs & Education category (United States searches only, 100 means peak popularity of the compared terms, 50 means half as popular, etc.).

The proposed curriculum leverages existing Computer Science and Engineering courses, with a single new course that needs to be developed. The CSE faculty have the necessary expertise to deliver the Data Science and Engineering curriculum – indeed the CSE department is already delivering a full complement of courses for the Master of Engineering degree in Data Science. Since the ongoing search for four new tenured/tenured track CSE faculty includes data visualization and natural language processing among the priority areas some of the new hires are also expected to contribute delivering the Data Science and Engineering curriculum. Undoubtedly, supporting an overall population of around 400 Data Science and Engineering majors will eventually require new faculty hires, additional IT and administrative staff, and commensurate increases in teaching assistant funding. These resources are expected to be allocated progressively as the program grows, supporting the need for additional course sections, student advising, and general IT and administrative support to accommodate the increased student enrollment and the new faculty hires.

Appendix: Core course descriptions and course schedule template

CSE 1010. Introduction to Computing for Engineers

Introduction to computing logic, algorithmic thinking, computing processes, a programming language and computing environment. Knowledge obtained in this course enables use of the computer as an instrument to solve computing problems. Representative problems from science, mathematics, and engineering will be solved.

2050. Data Structures and Object-Oriented Design

Prerequisites: CSE 1010

Introduction to fundamental data structures and algorithms. The emphasis is on understanding how to efficiently implement different data structures, communicate clearly about design decisions, and understand the relationships among implementations, design decisions, and the four pillars of object-oriented programming: abstraction, encapsulation, inheritance, and polymorphism.

2500. Introduction to Discrete Systems

Prerequisites: CSE 1010

Introduction to formal mathematical thinking including discrete systems and proofs. Discrete system topics include logic, set theory, basic number theory, basic combinatorics, functions, relations, sequences, sums, products, recurrence, and countability. Proof topics include direct proof, including proof by cases and induction, and indirect proof, including proof by contrapositive and contradiction.

CSE 2600. Introduction to Data Science and Engineering (new course)

Prerequisites: CSE 2050

Introduction to a broad selection of challenges and methodologies in working with big data. Topics to be covered include fundamental data science lifecycle topics such as data acquisition, management, integration, visualization, modeling, analysis, prediction, as well as data security, data privacy and ethics.

CSE 3000. Contemporary Issues in Computer Science and Engineering

Prerequisites: CSE 2050

Information management, the global and societal impact of computer science and engineering decisions, professional and ethical responsibility.

CSE 3140. Cybersecurity Lab

Prerequisites: CSE 2050.

Introduction to the design of secure systems. Explores issues that arise in multiple design phases to understand the limitations of the platform and the source of opportunities for attackers. Each unit will explore a system, its design, its vulnerabilities and how to exploit them, culminating with the creation, implementation and deployment of counter-measures to eliminate the vulnerabilities and nullify the threat.

CSE 3500. Algorithms and Complexity

Prerequisites: CSE 2050 and 2500

Design and analysis of efficient computer algorithms. Algorithm design techniques, including divide-and-conquer, dynamic programming, and greedy approaches. Graph

algorithms and advanced data structures. Worst-case and average-case analysis, reductions, and NP-completeness.

CSE 4502. Big Data Analytics

Prerequisites: CSE 3500 and MATH 2210Q

Focuses on basic concepts of data science and big data analytics. Different algorithmic techniques employed to process data will be discussed. Specific topics include: Parallel and out-of-core algorithms and data structures, rules mining, clustering algorithms, text mining, string algorithms, data reduction techniques, and learning algorithms. Applications such as motif search, k-locus association, k-mer counting, error correction, sequence assembly, genotype-phenotype correlations, etc. will be investigated.

CSE 4701. Principles of Databases

Prerequisites: CSE 3500

Fundamentals of data base design and data indexing techniques. Hierarchical, network, and relational data models. Data base design theory. Query languages, their implementation and optimization. Data base security and concurrent data base operations.

CSE 4820. Introduction to Machine Learning

Prerequisites: MATH 2210Q and STAT 3025Q

Recommended preparation: CSE 3500.

An introduction to the basic tools and techniques of machine learning, including models for both supervised and unsupervised learning, related optimization techniques, and methods for model validation. Topics include linear and logistic regression, SVM classification and regression, kernels, regularization, clustering, and on-line algorithms for regret minimization.

Data Science and Engineering Bachelor of Science Program

FRESHMAN YEAR

First Semester	Credits	Second Semester	Credits
Science lab course[1]	4	Science lab course[1]	4
MATH 1131Q – Calculus I	4	Math 1132Q – Calculus II	4
CSE 1010 – Intro Computing for Engineers	3	CSE 2050 – Data Structures & Object-Oriented Design	3
ENGR 1000 – Orientation to Engineering	1	ENGL 1007 – Seminar in Writing	4
Area 2 (Social Sciences)	3		
	15		15

SOPHOMORE YEAR

First Semester	Credits	Second Semester	Credits
Additional science course[1]	4	CSE 3140 – Cybersecurity Lab	2
CSE 2500 - Intro to Discrete Systems	3	CSE 3500 - Algorithms and Complexity	3
CSE 2600 – Intro to Data Science & Engineering	3	STAT 3025Q – Statistical Methods	3
MATH 2110Q – Multivariate Calculus	3	Area 2 (Social Science)	3
Area 1 (Arts and Humanities)	3	MATH 2210Q – Linear Algebra	3
	16		14

JUNIOR YEAR

First Semester	Credits	Second Semester	Credits
CSE 4701 – Principals of Databases	3	CSE 4502 – Big Data Analytics	3
CSE 4820 – Intro to Machine Learning	3	CSE 3000 - Contemporary Issues in CSE	1
Data Science & Engineering elective course 1	3	CSE Elective course 2	3
PHIL 1104 (Area 1) – Phil. and Soc Ethics	3	CSE Elective course 2	3
Elective	3	Area 4 (Diversity and Multiculturalism)	3
	15	Elective	3
			16

SENIOR YEAR

First Semester	Credits	Second Semester	Credits
CSE 4939W - Design Project I	3	CSE 4940 - Design Project II	3
Data Science & Engineering elective course 3	3	Data Science & Engineering elective course 4	3
Area 4 (Diversity and Multiculturalism)	3	Elective	3
Elective	3	Elective[6]	5
Elective	3		
	15		14

Additionally the program must include 1) one W course other than CSE 4939W, which may be used to satisfy other requirements or Free Electives, and 2) one E course of at least three credits in Environmental Literacy.

[1] A two-course sequence must be selected from one of the following sequences: CHEM 1127Q–1128Q; CHEM 1137Q–1138Q; CHEM 1147Q–1148Q; PHYS 1401Q–1402Q; PHYS 1501Q–1502Q; or PHYS 1601Q–1602Q. One additional science course must be selected from the the department not selected for the two-course sequence or from BIOL 1107, BIOL 1108, BIOL 1110, ERTH 1050, or ERTH 1051 and 1052

Three Degrees:

UCONN | COLLEGE OF LIBERAL ARTS AND SCIENCES

APPLIED DATA ANALYSIS BACHELOR OF ARTS PROGRAM

Focus on Algorithms in Social Science and Data Processing

Technical Skillset
Basic

Skills
Collect, Retrieve, Manipulate, Store, Analyze Data

Jobs
Data Scientists, Data Analysts

STATISICAL DATA SCIENCE BACHELOR OF SCIENCE PROGRAM

Focus on Mathematical Statistics

Technical Skillset
Intermediate

Skills
Collect, Retrieve, Manipulate, Store, Analyze Data

Jobs
Statisticians, Data Scientists, Data Analysts

UCONN | SCHOOL OF ENGINEERING

DATA SCIENCE AND ENGINEERING BACHELOR OF SCIENCE PROGRAM

Focus on Engineering Data from Machines that Translate to Industrial Systems

With an understanding of computer science, focus on computational proficiencies. All students will be required to complete a Senior Design project where they will design and implement solutions to real world data intensive problems

Technical Skillset
Advanced

Skills
Analysis of Complex Computing Problems, Design/Implement/ Evaluate Computing Based Solutions, Managing Complex Data Acquisition Systems

Jobs
Software Engineers, Business Analysts, Systems Engineers, Data Analysts, Electrical Engineers, DevOps Engineer, Java Developers

EXAMPLE
Pratt & Whitney = Jet Engines which produce data every second, resulting in sensors collecting data such as signal processing and artificial intelligence
VS
Financial (CLAS BS), and socio-economic/socially dynamic data (CLAS BA)