ILLINOIS TECH | College of Computing

Date: Jan 20th 2022

To: Undergraduate Studies Committee

From: Department of Applied Mathematics

Subj: Proposal for Bachelor of Science in Data Science

Attached is a proposal for Bachelor of Science in Data Science. This proposal has been approved by the Department of Applied Mathematics Undergraduate Studies Committee.

Proposal for B.S. in Data Science in College of Computing

Director and Chair of Steering Committee: Lulu Kang (AMAT)

Steering Committee: Robert Ellis (AMAT), Michael Pelsmajer (AMAT), Ioan Raicu (CS), Binghui Wang (CS), Kai Shu (CS)

1. Data Science Profession and Education

Data science has become an increasingly popular profession in industries, businesses, and government organizations across the economy and society, with a high projected job growth rate through 2030 and attractive median salary.[1] Employees trained with the analytical and computational knowledge and skills of data science are highly sought after. As one of the highest-paying professions, data science and related majors are also increasingly popular among students.

Although data science is related to traditional majors such as computer science, mathematics, statistics, and operations research; the multidisciplinary nature of data science requires a separate major for an efficient and comprehensive course of study. The curriculum must equip students with core skills in computer science, mathematics, and statistics; augmented by modern training in methodologies for data-driven analysis. This is rounded out by ethics and professional preparation, and yet still allows students to pursue methodological depth and breadth as well as broad applications in other disciplines.

In the last decade, data science education has been mainly at the graduate level. Many universities have opened data science master's programs or programs of similar names. Illinois Tech also opened MAS in Data Science in 2013 and MAS in AI in 2019 to meet the demands from students. Although these programs have been successful with graduates working in many reputable companies and industries, many universities have now also realized the importance and benefits of data science education at the undergraduate levels. Prominent examples include the UC Berkeley BA in Data Science, Carnegie Mellon University B.S. in Statistics and Machine Learning, and the brand new University of Chicago BA in Data Science.

[1] "Data Scientist" is the #2 job in the 2021 Glassdoor ranking as of October, 2021; with closely related jobs "Data Engineer" and "Machine Learning Engineer" also making the top 50. The US Bureau of Labor Statistics predicts 31.4% cumulative growth in data science and other mathematical science occupations between 2020-2030, with median salary of $98,230.

2. B.S. in Data Science in the College of Computing

Illinois Tech has been proud for being #1 in Illinois and #32 in the nation for lifting students from families in the bottom 20% of income to the top 20%. Since data scientists and related jobs are such high-paying ones, it would be a great career option for students from low-income families. Illinois Tech should definitely have its own undergraduate data science program. The newly founded College of Computing, whose purpose is to serve the education and research needs related to computing of the Illinois Tech community, has the responsibility and the capability to open such a program. The proposed B.S. in Data Science is under the administration of the CoC and is a truly interdisciplinary major across CS, Applied Mathematics, ITM, but also with significant opportunities for breadth through Humanities, Social Sciences, and engineering departments which are all outside of the CoC.

3. Program Requirements: Major: BS in Data Science

The curriculum of Data Science major focuses on the following aspects

- Mathematical foundations for data science;
- Computer Science algorithms, data structures, and programming;
- Data Science professionalism, communication, and ethics;
- Core data science knowledge in depth;
- Elective courses to broaden the horizon of students' knowledge in data science and related disciplines; and
- Hands-on learning experiences by development of complete solutions to realistic problems with real datasets.

The curriculum is both comprehensive and flexible. It emphasizes not only the knowledge and skills of data science, but also the professionalism, communication and ethics of the data science field, which is just as crucial to the success of the future careers of the students. The curriculum also allows students to take a minor in another discipline. A minor can be very helpful since the data science professional by nature is application-oriented and knowing other disciplines can only help a student to be more knowledgeable as a practitioner.

Data Science Requirements			**24–25**
DS 100		Introduction to the Profession	2
DS 151		Introduction to Data Science	4
One of the two options	DS 2XX and 3XX	Mathematical Foundations for Data Science I and II	3 + 3 or 4 + 3
	MATH 252 and (MATH 350 or MATH 380)	Introduction to Differential Equations **and** (Introduction to Computational Mathematics **or** Introduction to Mathematical Modeling)	
DS 251		Ethics and Privacy in Data Science	3
DS 451 or CSP 571		Data Science Life Cycle Data Preparation and Analysis	3
MATH 474 or MATH 476[2]		Probability and Statistics or Statistics	3
DS 484/MATH 484[3] or CS 484 or CS 422		Introduction to Statistical Learning/Regression Introduction to Machine Learning Data Mining	3

Applied Mathematics Requirements	**17**

[2] MATH 475 Probability is a prerequisite for MATH 476 Statistics, so students taking MATH 476 must take both courses, and MATH 475 counts as a Data Science Technical Depth Elective or Data Science Elective. MATH 475 and 476 give a stronger foundation which is needed in order to take certain advanced topics courses.

[3] Math 484 Regression will be updated and replaced by DS 484 Introduction to Statistical Learning.

MATH 151	Calculus I	5
MATH 152	Calculus II	5
MATH 251	Multivariate & Vector Calculus	4
MATH 332	Elementary Linear Algebra	3

Computer Science Requirements			**10–12**
One of the two options	CS 115 & CS 116	Object-Oriented Programming I and Object-Oriented Programming II	2 + 2
	CS 104 & CS 201	Intro to Comp Prgrm for Engrs and Accelerated Intro to Cmptr Sci	2 + 4
CS 331		Data Structures and Algorithms	3
CS 425		Database Organization	3

Communication: Select 1 of the following:		**3**
COM 421	Technical Communication	3
COM 423	Communication in the Workplace	3
COM 428	Verbal and Visual Communication	3
	Communications for the Workplace	3
INTM 301	Communication in the Workplace	3
ITM 300	Public Engagement for Scientists	3
SCI 522		

Ethics and Society: Select 1 of the following:		**3**
HIST 385	Women in Computing History	3
ITMM 485	Legal and Ethical Issues in IT	3
PHIL 374	Ethics in Computer Science	3
PHIL 375	Computer Ethics	3
PHIL 381	Artificial Intelligence, Philosophy and Ethics	3
SOC 362	Technology and Social Change	3
Data Science Technical Depth: Select 2–4 of the following:		**6–12**
MATH 435	Linear Optimization	3
MATH 475	Probability	3
MATH 476	Statistics	3
MATH 446/546	Introduction to Time Series	3
CS 429	Information Retrieval	3
CS 430	Introduction to Algorithms	3
CS 451	Parallel/Distributed Computing	3
CS 481	Intllgnc Txt Analys Knwldg Mgm	3
CS 422	Data Mining	3
CS 522	Advanced Data Mining	3
CS 577	Deep Learning	3

CS 584	Machine Learning	3
CSP 554	Big Data Technologies	3
MATH 535	Optimization I	3
MATH 563	Mathematical Statistics	3
MATH 564	Applied Statistics	3
MATH 569	Statistical Learning	3
MATH 574	Bayesian Computational Statistics	3

Data Science Electives		**6–12**
Any Data Science Technical Depth course listed above		
DS 4XX	Data Science Internship	3–6
MATH 497	Problem-Solving Projects in Business, Government, and Industry	3
COM 383	Social Networks	3
SSCI 325	Intermediate Geographic Information Systems	3
SSCI 480	Introduction to Survey Methodology	3
STAT 225	Introductory Statistics	3
CS 458 or ECE 443	Intro to Information Security or Intro Computer Cyber Security	3

CS 480	Introduction to AI	3
CS 487	Software Engineering	3
MATH 380	Intro to Mathematical Modeling	3
MATH 483	Design and Analysis of Experiment	3
CS 512	Computer Vision	3
CS 520	Data Integration, Warehousing, and Provenance	3
CS 578	Interact/Trans Mach Learning	3
CS 583	Probabilistic Graphical Models	3
CS 585	Natural Language Processing	3
CS 579	Online Social Network Analysis	3
CS 546	Parallel and Distributed Proc	3
CS 553	Cloud Computing	3
CS 554	Data-Intensive Computing	3
ECE 442	Internet of Things/Cyber Phys	3
ECE 510	Internet of Things/Cyber Phys	3
MATH 527	Machine Learning for Finance	3
MATH 565	Monte Carlo Methods in Finance	3
ITMS 418	Coding Security	3

ITMS 448	Cyber Security Technologies	3
ITMS 478	Cyber Security Management	3
ECE 308	Signals and Systems	
ECE 447	Artificial Intelligence and Edge Computing	3
ECE 501	Artificial Intelligence and Edge Computing	3
ECE 449	Object-Oriented Programming and Machine Learning	3
ECE 481	Image Processing	3
ECE 511	Analysis of Random Signals	3
ECE 520	Information Theory and Applications	3
ECE 521	Quantum Electronics	3
ECE 563	Artificial Intelligence in Smart Grid	3
ECE 565	Computer Vision and Image Processing	3
ECE 566	Machine and Deep Learning	3
ECE 567	Statistical Learning Processing	3
EMGT 363	Creativity, Inventions, and Entrepreneurship for Engineers and Scientists	3
EMGT 406	Entrepreneurship and Intellectual Property Management	3

SSCI 381	Computational Social Science	3

Minor		**15**
Any minor except Data Science	(advising toward subject matter expertise)	15 credits

(Minor is optional. If the student does not wish to take a minor, then they must take 6 extra credits of Data Science Technical Depth, 6 extra credits of Data Science Electives, and have 3 extra credits of Free Electives.)

Science Requirement and Electives		**10**
Natural Science and Engineering Requirements	See Illinois Tech Core Curriculum, section D	10 credits

Humanities and Social Science Requirements	**21**
See Illinois Tech Core Curriculum, sections B and C	21 credits
Interprofessional Projects (IPRO)	**6**
See Illinois Tech Core Curriculum, section E	6 credits

Free Electives	**6–9**

Total Credit Hours	**127**

4. Sample Curriculum

Year 1			
Semester 1 (Fall)	**Credit Hours**	**Semester 2 (Spring)**	**Credit Hours**
DS 100	2	Ethics and Society	3
DS 151	4	MATH 152	5
MATH 151	5	CS 116	2
CS 115	2	Hum or Soc Sci	3
HUM 200	3	Free Elec	3
	16		**16**
Year 2			
Semester 3 (Fall)	**Credit Hours**	**Semester 4 (Spring)**	**Credit Hours**
DS Elec	3	DS 2XX	3
MATH 251	4	MATH 474	3
CS 331	3	Minor Elec	3
Science Elec	4	Science Elec	3
Soc Sci Elec	3	HUM 300+	3
	17		**15**
Year 3			
Semester 5 (Fall)	**Credit Hours**	**Semester 6 (Spring)**	**Credit Hours**
DS 251	3	DS 3XX	3
CS 484	3	DS Tech Depth	3

MATH 332	3	Communications	3
Minor Elec	3	Minor Elec	3
Soc Sci 300+	3	Soc Sci 300+	3
		Free Elec	3
	15		**18**
Year 4			
Semester 7 (Fall)	**Credit Hours**	**Semester 8 (Spring)**	**Credit Hours**
DS 451	3	DS 4XX Internship	3
DS Elec	3	DS Tech Depth	3
Minor Elec	3	Minor Elec	3
IPRO	3	IPRO	3
Science Elec	3	HUM 300+	3
	15		**15**
Total Credit Hours: 127			

5. Program Requirements: Minor in Data Science

The requirements for Minor in Data Science are very specific. These five listed courses are targeted to quickly equip students with the most essential skills and prepare them for data-science-related jobs or applications.

15 credits

1. DS 151 Introduction to Data Science
2. DS 251 Ethics and Privacy in Data Science
   Prerequisite: DS 2XX Mathematical Foundations for Data Science
3. MATH 474 or MATH 476
   Prerequisite: MATH 251 or MATH 475
4. DS 484 Introduction to Statistical Learning/MATH 484 Regression or CS 422 or CS 484

Prerequisite: MATH 474 or CS 331

5. DS 451 Data Science Life Cycle or CSP 571
   Prerequisites: DS 251Ethics and Privacy in Data Science and one of DS 484 Introduction to Statistical Learning/Math 484 Regression or CS 422 or CS 484

6. Students Advising

The program administration will work with the Applied Mathematics and Computer Science Departments to find suitable faculty and/or staff to serve as academic advisors to students in the program. Such advisors will need to be familiar with the curriculum the data science domain, and the university policy regarding the undergraduate study. The students are required to regularly get in touch with the academic advisor to report and receive feedbacks on their study, courses selection, and their academic performances.

7. Admission

The admission requirements of the B.S. in Data Science program are similar to other undergraduate programs of the College of Computing. The administration staff of the program will work in coordination with the Undergraduate Admissions Office on marketing, recruitment and other aspects of the admission process.

8. Program Administration

- **Program Director (faculty).** Responsible for the overall content of the curriculum and ensuring that needed courses are developed and taught, in conjunction with the steering committee and the Applied Mathematics and Computer Science Departments. Runs regular program assessments. Builds and maintains a network of corporate, non-profit, and government employers of data scientists; for the purposes of (1) Securing Math 4XX internships; (2) Developing case studies and projects for incorporation into data science courses; and (3) Updating curriculum content with the evolving skill sets required of data science practitioners. Assists post-graduation job placement of majors. Assists with fundraising efforts. Appointed by agreement of the Applied Mathematics and Computer Science Department chairs.
- **Program Coordinator (staff).** Tracks majors through graduation and beyond, to maintain an alumni network to support identification of internship and career opportunities for current majors as well as engagement and fundraising. Supports guest speakers and seminars. Processes logistics for data science internships, and supports sharing agreements between faculty and outside organizations for internships, case studies, and data-sharing. Facilitates compliance with IIT

regulations and employment law. Forms and coordinates vertically-integrated student data science pods, mixing new and experienced students for mentoring in academic success, conducting team projects, and preparing for careers. (This may be a part-time position initially until the program grows sufficiently.)

- **Steering Committee**. A group of at least six faculty members, including the program director, chosen in equal numbers from the Applied Mathematics and Computer Science Departments, appointed by the departments. The committee will be solely responsible to review and approve all curriculum revisions and policy matters related to the program. The committee will report to the department chairs of the Applied Mathematics and Computer Science Departments.
- **External Advisory Committee.** An external committee of experts in data science technology and application from industry, government, and non-profit organizations. They will critique curriculum design; recommend real-world content for methods, case studies, projects, and ethics and privacy concerns; advise on best practices for professional communication and professional training; and promote students' career development and placement in data science internships. Significant representation will come from professionals working with community stakeholders, including on initiatives of interest to underrepresented and/or under-served populations.

9. Close Ties with MAS in Data Science and Other Graduate Programs

The proposed B.S. in Data Science is closely tied to the MAS in Data Science Program, MAS in Artificial Intelligence, and other MS and MAS programs in the College of Computing.

Particularly, students in the B.S. in Data Science program would have finished all the necessary prerequisites of the MAS in Data Science program when they finish the first two years of study. Also, the students will be qualified to pursue the MAS Artificial Intelligence degree directly after they graduate. Students also have the options to pursue an accelerated master's program, and finish the BS in Data Science and MAS in Data Science in five years or combine the BS in Data Science with other relevant master's programs within or outside of the college.

One of the most common themes of both the BS and MAS in Data Science is their emphasis on the professional training, which is best presented in project-oriented courses. The MAS in Data Science requires students to take a 6-credit capstone project course, CSP 572 Data Science Practicum. In this course, students work on real-world projects provided by the industrial

partners of the programs. Students will be supervised by both their academic advisor and the industrial partners, who both give final evaluation of students' work. Similarly, the BS in Data Science also requires students to gain hands-on experience of using Data Science knowledge and skills to solve real world problems. For one, students need to take IPRO courses (6 credits) that are already offered by various departments. Besides, *MATH 497 Problem-Solving Projects in Business, Government, and Industry*, which provisionally provides IPRO credit at the time of this proposal, offers students step-by-step guidance on developing a real data science solution. We also give students the opportunity to do *DS 4XX Data Science Internship*. Students can work on a paid or unpaid data-science-related internship while being advised and supervised by their advisor. *DS 451 Data Science Life Cycle* is another course that systematically teaches students on how to formulate their data science project, collect data, and construct solutions.

Given this common theme on professional training, the two programs can share resources on the industrial projects and partner relationships with industrial partners.

The MAS in Data Science emphasizes communication and ethics training. Similarly, the BS in Data Science not only emphasizes the training in communication and humanities, but also puts more weight on ethics training in data science. We have required two courses (6 credits) in ethics, privacy and philosophy, with one of these, DS 251, being specifically data-science-oriented. The other is to be selected from the Ethics and Society requirement cluster of courses, and suggests further free electives or minor courses for students particularly interested in these topics.

10. New Courses

- DS 100, Introduction to Profession (for Data Science)
- DS 151. Introduction to Data Science

Similar to U.C. Berkeley's course Data 8, an NSF-supported curriculum component specifically designed to support course adoption by other universities. DS 151 intends to introduce the critical concepts and skills in statistical inference, machine learning, and computer programming, through hands-on analysis of real-world datasets from various fields. Simplified, open-source data analysis and visualization tools are available that form a stepping stone to industry-standard programming languages, which broadens accessibility of this course beyond programming aficionados. The course should be constructed to fulfill the Computer Science section of the Illinois Tech Core Curriculum (D.3).

- DS 2XX and 3XX. Mathematical Foundations for Data Science

Covers all the necessary mathematical tools from differential equations, numerical analysis and computation, optimization, etc., that are relevant to data science but are not covered by Calculus I, II and III. Adds key methodological elements of the data-driven analytical process, including: structuring data, models and tools for missing data, metrics for data quality, exploratory data analysis, data visualization, outlier-detection and handling, fitting statistical distributions, hypothesis formulation and testing, validation of select analytical methods, and issues for prediction and generalization. Although these topics are covered in some existing MATH courses, they are scattered throughout 3 – 4 different 300 and 400-level courses. Therefore, it is very necessary to create these two new courses that abstract the most relevant mathematical topics and deliver them in the context of data science methods and applications. Presentation of techniques is accompanied by application to real-world examples.

DS 2XX Prerequisite: DS 151 or MATH 225. DS 3XX Prerequisite: DS 2XX.

- DS 251 (or 2YY). Ethics and Privacy in Data Science

This course takes a technical approach to exploring societal issues of ethics, fairness, responsibility, and privacy related to the collection, use, and generalization of data. Technical measures of fairness and equity are presented, along with their potential benefits and shortcomings. Statistical and algorithmic techniques for providing data privacy are discussed and critiqued in real-world contexts. Case studies are used to expose breakdowns in the data science pipeline that have led to inequities or actual harm, requiring mitigation or refactoring of the pipeline. Specific focuses include the effect of uneven data collection or misuse of data generalization on underrepresented groups, and built-in algorithmic or statistical bias. Prerequisite: DS 2XX Mathematical Foundations for Data Science.

- DS 484. Introduction to Statistical Learning

This course is the simpler version of MATH 569 Statistical Learning or CS 584 Machine Learning. It will be updated from and replace MATH 484 Regression, modernizing and tailoring the latter course towards the data science audience. Prerequisite: MATH 474 or MATH 475+476

- DS 451. Data Science Life Cycle

Stages of a data science project from start to finish: obtaining data, exploring data, determining what questions the data can answer, exploratory analysis, ethical impacts analysis and mitigation, hypothesis (re-)formulation, in-depth analysis, validation, and reporting. Case studies extensively included. Designed to educate the DS student in the

typical project life-cycle stages required in the data science professions. Prerequisites: DS 251 Ethics and Privacy in Data Science and one of DS 484 Introduction to Statistical Learning/MATH 484 Regression or CS 484 or CS 422.

- DS 4XX. Data Science Internship

A 3-6 credit course which combines faculty supervision of academic content with an embedded data science internship at a participating employer. The internship is designed to be consistent with the data science life cycle and the corresponding course DS 451. The internship content is negotiated between the faculty supervisor and employer, and includes several reporting stages and/or mentoring meetings with the faculty supervisor to ensure adherence to academic guidelines. Credit hours are prorated to (1) The portion corresponding to faculty supervision, which are tuition credits paid by the student; and (2) The portion corresponding to employment experience, which we intend to be tuition-free and for which the employer pays the student at least minimum wage. A clear delineation of student contact with faculty vs. employer determines paid work time. Prerequisite: DS 451 Data Science Life Cycle.

APPENDIX: STUDENT ASSESSMENT PLAN

BACHELOR OF SCIENCE IN DATA SCIENCE

Student Outcomes

1. Assess data collection, modeling, analysis, visualization, and explanation needs in the context of a client's needs
2. Appropriately collect, clean, evaluate, and prepare data for exploration, modeling, and analysis
3. Design, implement, and evaluate relevant computational systems to address data science needs
4. Understand key statistical analysis methods and be able to choose and implement appropriate models for a given data analysis problem
5. Effectively derive and communicate useful insights from data, including through storytelling and visualization
6. Communicate effectively in a variety of professional contexts
7. Function effectively as a member or leader of a team engaged in activities appropriate to the discipline
8. Identify, analyze, and mitigate ethical, privacy, and data and algorithmic bias issues in practical data science contexts

Assessment Process

Before every term, all faculty teaching undergraduate courses in the department are informed of the course outcomes and program outcomes that apply to their course(s) (see below). They are asked to include the course outcomes in their syllabus, and to ensure coverage of material supporting both course and program outcomes.

Every year one of the core course groups (DS 451/CSP 571, CS425, MATH 474/475/476, DS 251) for assessment data collection. Each such course maps to a subset of the program outcomes (see below).

The faculty teaching the courses each fall and spring are asked to collect a random subset (about 10%) of student work (final exams or final projects) that includes content to assess each program outcome that maps to the course.

Every summer, a rubric is applied to the student work mapping to each program outcome. The threshold we use is: 70% of a random student sample tested within a one-year window should achieve 70% of the points possible for questions matching each program outcome. The scoring of the rubric is done by faculty not teaching the course.

An assessment report is created for each academic year and discussed by the Data Science Program Steering Committee and any recommended actions are decided, documented in meeting minutes, and conveyed to the relevant departments for implementation. The entire assessment process is managed by the program director and documented in yearly assessment reports. The assessment reports are submitted to the IIT Director of Student Learning Assessment for their feedback and permanent storage in Blackboard.

Mapping of Student Outcomes to Courses

Student Outcome	**Relevant Courses**
Assess data collection, modeling, analysis, visualization, and explanation needs in the context of a client's needs	DS 451 or CSP 571
Appropriately collect, clean, evaluate, and prepare data for exploration, modeling, and analysis	DS 451 or CSP 571
Design, implement, and evaluate relevant computational systems to address data science needs	CS 425 DS 451 or CSP 571
Understand key statistical and machine-learning analysis methods and be able to choose and implement appropriate models for a given data analysis problem	MATH 474, 475, 476 DS 484/MATH 484 or CS 484 or CS 422
Effectively derive and communicate useful insights from data, including through storytelling and visualization	DS 451
Communicate effectively in a variety of professional contexts	COM courses (how to assess?)
Function effectively as a member or leader of a team engaged in activities	DS 451 DS 4XX Data Science Internship

appropriate to the discipline	
Identify, analyze, and mitigate ethical, privacy, and data and algorithmic bias issues in practical data science contexts	DS 251 DS 451