Curriculum Committee Resolution 23/24-17

Date: May 8, 2024
To: St. Olaf Faculty
From: Curriculum Committee
Re: New Major - B.A. in Statistics and Data Science

The Curriculum Committee will move that the faculty approve the following new major.

Program Description:
With the growing abundance of data gathered in nearly every field, statistics and data science methods have become invaluable for transforming data into useful information. A statistics and data science (SDS) major at St. Olaf gives students the quantitative, analytical, and communication skills to navigate an increasingly data-rich world. As a subject, statistics and data science lies at the intersection of mathematics, computer science, and statistics and connects naturally to the sciences (natural and social), the humanities, and even the arts.

Students pursuing a statistics and data science major at St. Olaf will take courses that blend theoretical and practical concepts, exploring foundational ideas in computational thinking, statistical modeling, and mathematical underpinnings. SDS emphasizes a hands-on approach, and students will gain proficiency in programming languages and statistical software commonly used in the field. In addition, SDS courses encourage students to examine ethical considerations and societal implications of data collection and analysis, a particularly important consideration in the growing world of "big data".

As a part of the Department of Mathematics, Statistics, and Computer Science (MSCS) students that major in statistics and data science will have access to faculty and resources in these related but distinct disciplines. Faculty in SDS come from a variety of backgrounds, and use their applied research experience to guide students through collaborative projects and research opportunities that prepare students to contribute meaningfully to their chosen field.

Requirements for the Major:

Required Statistics & Data Science Courses (5 credits)		
MSCS 164 (SDS 164)	Data Science 1	1.00
MSCS 264 (SDS 264)	Data Science 2 (prereq MSCS 164)	1.00
STAT 172 (SDS 172)	Statistics 1	1.00
or ECON 260	Introductory Econometrics (or AP Statistics)	
STAT 272 (SDS 272)	Statistics 2 (prereq STAT 172 and MATH 120)	1.00
MSCS 341 (SDS 341)	Algorithms for Decision Making (prereq MSCS 164, STAT 272, and MATH 220) [1]	1.00
Required Mathematics Courses (2 credits)		
MATH 120	Calculus I	1.00
or MATH 119	Introduction to Calculus	
MATH 220	Elementary Linear Algebra	1.00
Required Philosophy Courses (1 credit)		
PHIL 244	Philosophy of Science [2]	1.00
or PHIL 251	Science, Ethics, and Religion	
Electives Courses (2 credits) - Including at least one 300-level course from Statistics or Data Science courses listed below.		
Statistics Depth Courses - select at least one of the following:		1.00
<u>Statistics Depth Level II Courses</u>		
STAT 282 (SDS 282)	Topics in Statistics (prereqs vary)	
STAT 284 (SDS 284)	Biostatistics: Design and Analysis (prereq STAT 272)	
ECON 384	Econometrics: Cross-Sectional and Panel Data	
ECON 385	Econometrics: Time Series and Forecasting	
SDS 2XX Spatial Data Analysis (prereqs MSCS 164 and STAT 172) [3]		
<u>Statistics Depth Level III Courses</u>		
STAT 316 (SDS 316)	Advanced Statistical Modeling (prereqs STAT 272 and MATH 220)	
STAT 322 (SDS 322)	Statistical Theory (prereqs STAT 272 and MATH 262)	
STAT 382 (SDS 382)	Advanced Topics in Statistics (prereqs vary)	
Data Science Depth Courses (1.0 credit) - select at least one of the following:		1.0
<u>Data Science Depth Level II Courses</u>		
BIO 315	Principles of Bioinformatics	
CSCI 379	Foundations of Artificial Intelligence	
SDS 250 Principles of Data Visualization (prereq MSCS 164 or STAT 172) [3]		

SDS 280 Topics in Data Science (prereqs vary) [3]	
Data Science Depth Level III Courses	
SDS 3XX High Dimensional Data Analysis (prereq MSCS 164 and MATH 220) [3]	
SDS 380 Topics in Data Science (prereqs vary) [3]	
TOTAL CREDITS	**10**

[1] *Prerequisites that are not yet proposed. Current prerequisites for MSCS 341 are MSCS 164 or MSCS 264 or permission of the instructor. It will be important to include "or permission of instructor" for next year, as there will be a transition period where not all current students switching to the new major will have the prerequisites. Incoming students will have the major available to them and we will be enforcing the new prerequisite structure on the course if approved.*

[2] *The first offering of PHIL 244 Philosophy of Science (with Data Science Module) that will satisfy the SDS major will be in 2025-26.*

[3] *Courses that are not yet proposed. All except SDS 280 and SDS 380 have been taught as topics courses previously.*

Rationale:

Data Science is an emerging discipline that lies at the intersection of Statistics, Computer Science, and Mathematics (MSCS). It can be defined as the science of extracting information from data to uncover patterns, find relationships, draw conclusions, and make decisions. Due to advances in computing and the ubiquity of data, the way in which data is leveraged to understand complex problems has changed dramatically over the past ten to twenty years. A modern data scientist or statistician must be extremely flexible in framing questions and applying methods to derive knowledge from large, unstructured, and observational data. They must understand the entire data pipeline, from study design and data collection, to tidying and organization, to modeling and prediction, to communication. Navigating that pipeline requires statistical fundamentals, mathematical principles, and computing acumen; written, oral, and visual communication skills; understanding of the domain area; and, application of ethical reasoning. With a nationally-recognized department where mathematics, statistics, and computer science happily co-exist, surrounded by excellence in the natural sciences, social sciences, and the arts and humanities, St. Olaf is well-positioned to be a leader among liberal arts undergraduate Statistics and Data Science programs, although many institutions have gotten a jump start by adding their own majors over the past decade.

Rather than simply a major in "Data Science", we are proposing a major in "Statistics and Data Science" to indicate a statistics-flavored data science program which reflects the historic strengths of our department. We believe that "Statistics and Data Science" sends a meaningful signal to prospective and current students (as well as future employers) about intended learning outcomes from the major. This also reflects a movement beyond "Statistics" to place a broad emphasis on the entire data pipeline. If Data Science is represented as a continuum from mathematical statistics and modeling on one end, to heavy computational algorithm-writing and data-warehousing on the other end, our proposed major would land closer to the statistics end,

with an increased emphasis on the computing acumen required to acquire and organize data and to run statistical learning algorithms. While Statistics is more "Here is a nice, clean spreadsheet of data; what does that tell us about our pre-defined research question?", Data Science is closer to "Here is a bunch of messy data (maybe even text and images); what can we learn from it?" The skill sets required to answer those questions are nicely complementary.

The groundwork for a major in Statistics and Data Science began being laid during the Strategic Resource Allocation Project in Spring 2019. While details in the proposed major have morphed, our 5-year plan (extended due to the 6-to-5 transition) to pilot courses, bring in external (p)reviewers, and engage in conversations with constituencies across the college targeted the submission of a proposal for a new major this academic year. In particular, a team of 3 nationally-respected (p)reviewers produced a report in March 2022 that strongly endorsed our plan for a new major in Statistics and Data Science. Part of our desire for a Statistics and Data Science major stems from the growing demand for courses in this area, both locally and nationally; in many ways, a new major would formalize an academic path that many students have been piecing together for themselves anyway, while making sure key fundamentals in mathematics, computing, and ethical reasoning are not missed.

Over the past decade, the Statistics and Data Science concentration (formerly the Statistics concentration) has grown from just under 30 students per year to the largest concentration on campus with over 80 students per year. There is no one major that comprises the majority of current concentrators; students from across the college are expressing a strong interest in anything related to data science, observing that undergraduate experiences with modern statistical methodologies and data science are "door openers" for careers in data-intensive professions and graduate programs in a growing number of disciplines. There are important questions across a seemingly endless array of fields that students can help address with an understanding of statistics and data science—from treating disease to global warming to insights on social issues such as income inequality and mass incarceration. The admissions staff reports that an increasing number of prospective students are aware of data science and asking about St. Olaf's program, and the Piper Center is encouraging as many students as possible to acquire data science skills. A Statistics and Data Science major would provide a disciplinary home for those students, while good advising could help non-majors find 100- and 200-level courses to prepare them for analytical contributions in their primary domain of interest.

Having a Statistics and Data Science major will allow MSCS, through programs such as the Center for Interdisciplinary Research (CIR), to expand its ability to collaborate with members of our faculty whose research interests have strong data science components and do not naturally fit within the structures of traditional statistics methodology, while also making our students more competitive for external internships early in their academic careers. Significantly, a Statistics and Data Science major has strong potential to increase STEM diversity. For example, the proportion of women in statistics has traditionally been higher than in other STEM fields. In addition, a Statistics and Data Science major will provide additional entry points into quantitative fields for students from diverse backgrounds, especially for those without the traditional calculus sequence who are nevertheless motivated to contribute to quantitative aspects of relevant and

often socially critical questions. Since data science is rarely taught at the high school level today, a course like Data Science 1 will be a level playing field for all students considering the journey.

We will be sunsetting the concentration with the introduction of the Major. Additionally, it is our understanding that the college would not support an SDS major and a concentration in the same subject (which makes sense). We are sad to see the concentration be discontinued as we know that has been a great opportunity for many students. We see the major as the more appropriate application of our resources. Additionally, we'll be able to enforce a prerequisite structure that will allow us to better scaffold the curriculum.

Students who currently have a declared SDS concentration will be encouraged to complete their concentrations. Current students with anticipated graduation years of 2025-2027 will continue with the concentration. The concentration will end in spring 2027 as the last SDS concentration students graduate. The concentration will not be available to students entering after the 2023-24 academic year. If a current student inquires about the new major, they will work with the program director and meet with their advisor to discuss feasibility.

Intended Learning Outcomes:
The Statistics and Data Science major is an integrated and interconnected set of courses, reflecting the interdisciplinary nature of the field and welcoming students from all backgrounds and experiences. Students will be grounded in foundational ideas from mathematics, such as probability, linear algebra, optimization, and multivariate thinking, alongside foundational ideas from computer science, such as algorithmic thinking, abstraction, project workflow, and reproducibility. Students will be immersed in collaborative team settings common to professional data scientists. In addition, students will acquire depth in their understanding of both statistics and data science, while being constantly challenged to consider ethical issues in their work and to apply statistics and data science principles to domains of expertise.

Students will demonstrate the ability to:
1. Acquire and curate data of all types.
2. Perform exploratory data analyses through data visualization and numerical summarization.
3. Build, assess, and interpret machine-learning and statistical models.
4. Communicate findings effectively and responsibly to a variety of audiences.
5. Apply analytical thinking to formulate problems, plan data collection/acquisition and engage in the data analysis process to provide insights in an integrated project.
6. Identify and critique multiple perspectives regarding data ethics.