Major in Mathematics - Mathematical Data Science

1 Major Description

This undergraduate major is designed to help prepare Math/Stat students for careers in Data Science, a field rapidly expanding in the private sector.

Because computer programming and machine learning are the primary components of Data Science, most data scientists currently have been trained as computer scientists. Our proposed major will cover a core of programming courses, but complemented by Math and Statistics courses. The training in abstraction and reasoning, and the deep understanding of the relevant algorithms, will make our graduates extremely useful members of data science teams working in the private sector.

2 Rationale for new major: types of students, expected number of students and documentation, community need and relevance

The field of Data Science is very rapidly expanding in both academia and the private sector, and promises to be a major source of high paying jobs for years to come. See

https://www.omnesgroup.com/in-demand-tech-jobs-for-2022/,

https://www.simplilearn.com/highest-paying-tech-jobs-article

The field, roughly speaking, is a set of techniques of detecting patterns in very large data sets. Although the primary tools, such as Machine Learning, are computer science based, together with knowledge in the fields associated to the data sets (for instance knowledge of biology), a vital role is played by mathematics and statistics.

The proposed major will give its graduates a solid core in programming skills, but also superior knowledge in math/stat courses such as Linear Algebra and Statistics for Data Science. This specific knowledge, plus the superior abstraction skills that come with taking a Math/Stats degree, will make them valuable members of teams working on Data Science in the private sector.

We expect our initial cohort to be small, and to graduate 10 students per year as we establish the program. Once we have established a solid job placement record for our graduates, we expect to double or triple this number. We expect the student pool to consist of some portion of our current majors, and other mathematically gifted students who will think this new major will lead to an interesting, well-paid job.

The need and relevance for this major, both locally and nationally, should be clear from the national hiring figures. Locally, our graduates should be able to contribute to the medical,

business, financial, and aerospatial industries, among others. The unique skills of our majors will make them highly sought after in the job market.

3 Impact of this major on other units

This program will be targeting students who are primarily interested in the study of mathematics/statistics, and should not have a major impact on student numbers in other units.

The degree will have some impact on the School of Computer Science because we require several CS courses. Since those courses are already part of their offerings, we do not expect our program to put any major demands on their resources. The CS School is supportive of our degree, and they agree to include their five courses in our program requirements; we have included in our package a letter from their chair to that effect. We expect our majors and theirs to have some informal collaboration to the benefit of all.

4 Required courses

4.1 Required Math/Stat courses

MAC 2311	Calculus I	4
MAC 2312	Calculus II	4
MAC 2313	Calculus III	4
MAP 2302	Differential Equations	3
MAS 3105	Linear Algebra I	3
MAS 4XXX	Linear Algebra II	3
MAS 4$XXXL$	Linear Algebra II Lab	1
MAD 2104	Discrete Mathematics	3
STA 4321	Mathematical Statistics I	3
STA 4322	Mathematical Statistics II	3
STA 4XXX	Statistical Data Science	3
MAP4XXX	Optimization	3
MAP4$XXXL$	Optimization Lab	1
STA 4XXX	Statistical Machine Learning	3
STA 4$XXXL$	Statistical Machine Learning Lab	3
MAT 4XXX	Senior Design Project	3

4.2 Required Courses Comp

COP 2210	Computer Programming I	4
CAP 2752	Fundamentals of Data Science	3
COP 3337	Computer Programming II	3
COP 3530	Data Structures	3
COP 4710	Database Management	3

4.3 Completion Requirements

Year 1 and 2

MAC2311	Calculus I	4
MAC2312	Calculus II	4
MAC2313	Calculus III	4
MAD 2104	Discrete Mathematics	3
COP 2210	Computer Programming I	4
CAP2752	Fundamentals of Data Science	3
MAS3105	Linear Algebra I	3

Year 3 Fall

MAP2302	Ordinary Differential Equations	3
COP 3337	Computer Programming II	3
STA 4321	Mathematical Statistics I	3

Year 3 Spring

COP 3530	Data Structures	3
MAS $4XXX + L$	Linear Algebra II and Lab	$3 + 1$
STA 4322	Mathematical Statistics II	3

Year 4 Fall

MAP $XXXX + L$	Optimization and Lab	$3 + 1$
STA $4XXX$	Statistical Data Science	3
COP 4710	Database Management	3

Year 4 Spring

STA $4XXX + L$	Statistical Machine Learning and Lab	$3 + 1$
MAP $4XXX$	Senior Design Project	3

Note: +1 indicates a lab component

4.4 New Courses

MAS 4XXX. **Linear Algebra 2 (3+1)** Matrix decomposition, orthogonal projection, QR factorization, diagonalization of a matrix using eigenvalues and eigenvectors, Cholesky decomposition, singular value decomposition. Coding for implementation. With labs. Prerequisites: MAS 3105.

MAP 4XXX. **Optimization (3+1)** Unconstrained optimization, linear optimization, nonlinear constrained optimization, convex analysis and optimization, determministic and stochastic dynamical optimization. With labs. Prerequisites: Linear Algebra 2 (MAS 4XXX) and MAC 2313.

STA 4XXX Statistical Data Science (3). Linear and Multiple Regression. Logistic Regression. Bootstrap and resampling. Penalized regression (LASSO/ridge/elastic net/splines). Gaussian processes. Multiple testing. Bayesian methods. Time Series Analysis. Pre-requisites: STA 4322.

STA 4XXX Statistical Machine Learning (3+1). Precision/Recall. Decision Trees. Bagging/Boosting. Linear Discriminant analysis. Support Vector Machines. Clustering. Principal Components Analysis. Neural networks and Deep Learning. Matrix completion. Reinforcement Learning/Markov decision process. With labs. Pre-requisites: Statistical Data Science-STA 4XXX, Optimization-MAP 4XXX and COP 3337.

MAP 4XXX Senior Design Project (3). Capstone course includes:Project statement, data cleaning, applying machine learning to cleaned data, written report and oral presentation.

Pre-requisites: Statistical Machine Learning (STA 4XXX), Database Management (COP 4710), Optimization (MAP 4XXX).

5 Compliance with the BOG Regulation 8.011 Authorization of New Academic Degree Programs and Other Curricular Offerings.

Our major complies with the BOG Regulation 8.011 Authorization of New Academic Degree Programs and Other Curricular Offerings.

6 Majors within a degree must share a minimum of 15 percent of common core or prerequisite courses to comply with BOG policy on degree structures.

The following courses are requirements common to both the proposed major and to our current Math major: MAC 2311 (4 credits), MAC 2312 (4 credits), MAC 2313 (4 credits), MAC 3105 (3 credits), MAP 2302 (3 credits), STA 4321 (3 credits). This total of 21 credits is sufficient to meet the state guidelines of 15 percent overlap.

7 Resources required: faculty, library, facilities including laboratories

Resources required for this major include:

1- 5 new courses will be offered per year, and three of them come with one credit labs. We will use our current pool of TAs to generate the necessary resources: some of our TAs are used

as graders, etc. For this new program, they can staff the lab courses, and teach lower division classes, thereby freeing faculty to teach the upper division courses.

2- a dedicated lab space where the majors can interact and work on the lab component of their courses. We anticipate using DM 409A in the evenings for this. This room is already controlled by the Math Department, and is usually empty after 6PM. The actual lab work will be done on the students' laptops, so students won't need any resources for that.

3- The labs will have a dedicated TA in attendance, who will be selected from our current group.

8 Budget requirements; majors that require budget need a budget justification

The program needs no new resources, so no budget necessary.

9 Proposed catalog copy

Major in Mathematical Data Science

This undergraduate major is designed to help prepare Math/Stat students for careers in Data Science, a field rapidly expanding in the private sector.

Because computer programming and machine learning are the primary components of Data Science, most data scientists currently have been trained as computer scientists. Our proposed major will cover a core of programming courses, but complemented by Math and Statistics courses that will make our graduate a useful complementary part of a data science team working in the private sector.

Courses required for degree:

Required courses from Mathematics and Statistics

MAS 3105	Linear Algebra I	3
MAS 4XXX	Linear Algebra II	3
MAS 4$XXXL$	Linear Algebra II Lab	1
MAD 2104	Discrete Mathematics	3
STA 4321	Mathematical Statistics I	3
STA 4322	Mathematical Statistics II	3
STA 4XXX	Statistical Data Science	3
MAP 4XXX	Optimization	3
MAP 4$XXXL$	Optimization Lab	1
STA 4XXX	Statistical Machine Learning	3
STA 4$XXXL$	Statistical Machine Learning Lab	1
MAP 4XXX	Senior Design Project	3

Required Courses from Computer Science

COP 2210	Computer Programming I	4
CAP 2752	Fundamentals of Data Science	3
COP 3337	Computer Programming II	3
COP 3530	Data Structures	3
COP 4710	Database Management	3