College of Liberal Arts and Sciences
Office of the Dean
University of Iowa
240 Schaeffer Hall
Iowa City, Iowa 52242-1409
Phone: 319-335-2625 | Fax: 319-335-3755
Email: clas@uiowa.edu | Web: clas.uiowa.edu

March 26, 2021

John C. Keller
Associate Provost for Graduate Education
Dean of the Graduate College
Graduate College
201 Gilmore Hall
The University of Iowa

Dear Dean Keller:

The College of Liberal Arts and Sciences is very pleased to endorse the proposal for a MS in Data Science to be housed in the Department of Statistics and Actuarial Science in the College of Liberal Arts and Sciences. The College's endorsement carries with it a provision that the target enrollments and projected completion rates be assessed in 2024 and 2026 and the necessary modifications or elimination of the degree program be made as a result.

The proposed MS in Data Science offers an avenue through which graduate students can receive the advanced professional training in the technical, methodological and analytical competencies specific to the growing number of career opportunities for data scientists in the health sciences, industry, and government agencies.

As the proposal notes, no additional resources will be necessary to support the program. Several courses currently offered by the Department of Statistics and Actuarial Sciences will be revised to conform to the needs of the new degree and one new course will be added; these modifications and additions will be made to the curriculum by the current faculty and no new faculty will be required in order to deliver them. The Department of Statistics and Actuarial Science anticipates that students earning the BS in Mathematics, Computer Science, Statistics, Actuarial Science, Physics, Chemistry, and Biology will matriculate to the MS in Data Science and envisions possibilities for developing a U2G option for interested undergraduate students.

Sincerely,

Sara Sanders, Ph.D, MSW
Dean and Director of Diversity, Equity, and Inclusion
UI Alumni Association Dean's Chair
in the Liberal Arts and Sciences

C: Christine Getz, Kung-Sik Chan

Proposal to Create Master of Science in Data Science

Program College: Graduate College
Curricular College: Liberal Arts and Sciences
Academic Unit: Statistics and Actuarial Science
Degree Objective (with and/or without thesis): Master of Science without thesis
Program: Data Science
Subprogram: N/A
Offers a Certificate Program: N/A
Available to: degree-seeking students
Advisor: Jian Huang
Effective Session: Fall 2022
Declarable in ISIS: No
CIP Code (30.7001) – Data Science
Course Prefix: DATA
Distance Ed: On Campus

Table of contents:

1. Purpose of the program

The Department of Statistics and Actuarial Science proposes a new graduate program, Master of Science in Data Science (MSDS).

The goal of the proposed MSDS program is to train data scientists who have the analytical and technical skills to explore, formulate and solve complex data-driven

problems in science, industry, business and government. The program focuses on the theory, methodology, and application of techniques for working with and learning from data. With emphasis on statistical thinking and foundations, the MSDS program rigorously prepares its graduates with the abilities to develop and implement new or special-purpose analysis and visualization tools. The program will also promote a fundamental understanding of how to quantify uncertainty when making data-driven decisions. The program will prepare graduates interested in academic, industrial, business or government positions that involve data visualization and modeling, managing reproducible data analysis workflows, and collaborating with scientists and other data stakeholders.

The Department of Statistics and Actuarial Science is well positioned to develop the MSDS program. The Department has a strong data science group focusing on several core areas of data science, including Bayesian statistics, computational statistics, machine learning, statistical learning, and visualization and data technologies in addition to our globally visible expertise in actuarial science, categorical data, computing, high dimensional statistics, time series, spatial statistics, stochastic processes and survival analysis. The Department offers undergraduate majors and graduate programs in statistics and actuarial science, and partners with the Department of Computer Science to offer the BS program in data science. In addition, the Department collaborates with the Departments of Geographical and Sustainability Science, Political Science, and Sociology and Criminology to offer the undergraduate Certificate in Social Science Analytics. Therefore, the Department has the required faculty expertise and programmatic focus to offer the proposed MSDS program.

Furthermore, the proposed MSDS program is logically situated between our current BS Data Science (BSDS) program and PhD Statistics program with data science track. The MSDS will naturally be an attractive U2G option for the BSDS students to obtain a BS degree and an MS degree in 5 years. Therefore, we expect that the proposed MSDS program will help make our existing BSDS program more attractive to students who are interested in obtaining an MS degree in data science at a faster pace than the standard route to an MS degree. We will seek the approval for such an option as soon as the MSDS program is approved. The MSDS can also serve as a pathway to our PhD Statistics program with data science track. In particular, this should be helpful to the MSDS students who are interested in pursuing a PhD and completing the degree requirement in a relatively shorter period comparing with those who enter the PhD program from a different track, since some required courses for the PhD Statistics with data science are also required in MSDS, including Statistical Learning (STAT: 4540), Data Visualization and Data Technologies (STAT: 4580) and Statistical Computing (STAT: 5400).

It will also be mutually beneficial to our strong undergraduate and graduate programs in Actuarial Science. The Department of Statistics and Actuarial Science has been

designated as a Center of Actuarial Excellence (CAE) since 2009; it is a distinction shared by only 16 major research universities worldwide. The insurance industry has been expanding the uses of data in directing their business; consequently, predictive analytics emerges as one of the predominant topics in professional examinations and actuarial science research.

2. Need and Demand

Data of various types, complexity and sizes (especially complex, massive data) are increasingly collected and exploited across fields, widening the integration of data directed applications in society, science, business, etc. The age of massive data is unfolding in real time, reshaping business, industrial, and medical practices and thereby fascinating the public and changing their daily lives, while challenging both academia and industries to develop novel approaches to learning with all sorts of data. It is, therefore, an opportune moment for the Department of Statistics and Actuarial Science to propose and develop a MS Data Science program.

Companies and organizations of every size and industry – from Google, Amazon, medical research centers, pharmaceutical companies, government agencies to retail stores – are looking for experts to help them deal with data. The growing interest and demand for data scientists is largely driven by the convergence of two factors. First, we collect an unprecedented amount of data, from internet browsing logs, mobile telephony, medical imaging, genetic sequencing, and even personal information such as number of steps taken or heart rates over time. The second factor driving this sector is the emergence of new computationally efficient and easy-to-use software tools that can extract useful patterns from this vast morass of raw data. These tools are the result of fundamental developments in statistics, computer science and machine learning which, when coupled with technological advances in the underlying hardware, lead to new computational techniques capable of analyzing extremely large and unusually complex data sets.

Almost every scientific research project (quantitative analysis, modern business operation, monitoring by government agencies such as the Center for Disease Control that tracks the Covid-19 pandemic) produces, collects and mines large amount of data. Success very much depends on a ready supply of technologically literate employees able to perform sophisticated data analysis. As demand for new and better technology grows, demand for data scientists will grow as well. Also, as data structure becomes more complex and data volume becomes increasingly large, requirements for data

science positions and skills have become more demanding. According to the Bureau of Labor Statistics (BLS), an MS degree is typically required to enter this field[1] .

To meet the enormous demand for data scientists in academia, industry, business and government, higher educational institutions have been racing to establish data science and related programs to train qualified graduates in recent years. Among the Big Ten Academic Alliance (BTAA) institutions, eight have BS and/or MS programs in data science, six have business-oriented MA or MBA programs in business analytics. Nationally, at least nine institutions have both data science (MPS data science, MS data science and MS Statistics/Data Science options, residential) and business analytics (MS business or MBA with business analytics concentration, residential or online). Also, among BTAA institutions and nationally, at least five institutions have both residential business analytics (under various names) and data science MS programs. In the State of Iowa, the Iowa State University's Debbie and Jerry College of Business offers the Master of Business Analytics program[2], an online program for employed professionals that addresses the challenges of dealing with data analytics and business intelligence. It differs from the proposed MSDS program which is a residential, generalist degree designed for students with a liberal arts BS degree. More detailed information about the existing data science and related programs in the nation is provided in Appendix A.3. The proposed MSDS program, if approved, will be the only MSDS program in the State of Iowa. It will help the University of Iowa remain competitive nationally in the broad field of data science education and research and offer more data science training opportunities to the students in Iowa and the nation.

These recently established data science programs by higher education institutions demonstrate that there is a growing market and high demand for the graduates trained in data science. For instance, the MSDS program at the University of Minnesota has enjoyed robust growth since its inception in 2015, seeing steadily rising annual enrollment numbers: 53, 48, 59, 71, 86, 106, with an admission rate around 20%, which is much lower than their overall graduate student admission rate of about 45%[3]. The Statistics MS Data Science program in the University of Wisconsin-Madison has experienced similar, strong growth since its inception in 2015, with increasing annual

[1] https://www.bls.gov/ooh/computer-and-information-technology/computer-andinformation-research-scientists.htm#tab-1

[2] https://www.ivybusiness.iastate.edu/masters/business-analytics/

[3] https://apps.grad.umn.edu/stats/ad/1231300.shtml

enrollment numbers: 25, 43, 47, 60, 86 over the past five years (excluding year 2020) and the lowest admission rate at 47% in 2019-20[4][5].

However, the supply of well-trained graduates in data science is far from meeting the demand. Data science has been one of the fastest growing industries over the past few years, and with the ever-increasing demand for data, the field is poised to grow for years to come. Indeed, data science has only just begun to expand as a discipline into industries and government agencies other than technology. For example, in 2019, the US government's Office of Personnel Management (OPM) for the first time recognized the job title "Data Scientist" as an official role in government agencies[6]. And in June the same year, the Department of Education National Center for Education System (NCES) released the new Classification of Instructional Programs (CIP) Codes for data science, data analytics, and applied statistics. According to the Bureau of Labor and Statistics (BLS), employment growth of computer information and research scientists, which include data scientists, from 2019 to 2029 is 15%. PricewaterhouseCoopers (PwC) describes data science as among the most sought-after positions in America in their report on investing in America's data science and analytics talent[7]. Data science jobs ranked No. 3 on Glassdoor's list of best jobs in America 2020 with median base salary of $107,801 and job satisfaction level of 4 out of 5[8]. Data scientists have also been ranked as No. 3 in LinkedIn's Emerging Jobs Report 2020 with an estimated 37% annual growth rate[8]. Data from job boards like LinkedIn, Glassdoor and Dice.com consistently show strong job demand for data scientists. A search for "data scientists" (performed on February 13, 2021) on LinkedIn[9], Glassdoor[10], and dice.com[11] yielded 27,568, 29,354 and 27,545 data scientist jobs, respectively. Such demand exceeds the number of data scientists that can be trained in all the existing data science and related programs in higher education institutions.

---

[4] https://tableau.wisconsin.edu/views/ProgramProfilesGraduateSchoolExplorer/HomePage?Select%20Academic%20Plan%20(Universal%20Filter)=Statistics%20MS&Select%20Named%20Option%20(Universal%20Filter)=Data%20Science&:iid=

[5] &:isGuestRedirectFromVizportal=y&:embed=y

[6] https://www.chcoc.gov/content/data-scientist-titling-guidance

[7] https://www.pwc.com/us/en/library/data-science-and-analytics-skills.html

[8] https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm

[8] https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging Jobs Report_U.S. FINAL.pdf

[9] https://www.linkedin.com/jobs/data-scientist-jobs/

[10] https://www.glassdoor.com/Job/data-scientist-jobs-SRCH_KO0,14.htm

[11] https://www.dice.com/jobs/q-Data+Scientist-jobs

Finally, there is also strong interest in data science from undergraduate and graduate students across many subject fields because of the excellent employment and career opportunities. For example, there has been a sustained increase in the enrollment of students from many disciplines in two of our existing data science courses, namely, Statistical Learning (STAT: 4540) with enrollment numbers: 41, 26, 47, 57 and 62 over the period from fall 2016 to fall 2020, and Data Visualization and Data Technologies (STAT: 4580) with enrolment numbers: 17, 33, 38, 39, 48 from spring 2017 to spring 2021; over 65% of the students in these courses were from disciplines other than Statistics or Data Science. These data suggest that it is imperative to offer the proposed MSDS program to meet the student demand for data science training. They also indicate that the proposed MSDS program will be attractive to many students from Iowa, as well as the nation and the world.

3. Description of the program

We describe here the main academic characteristics of the proposed MSDS program. Additional curriculum details are provided in Appendices A.1 and A.2.

Target applicants: Students with a BS degree in any subject area who have taken and done well in calculus and linear algebra courses, and have some background in computer programming (e.g., python, R and/or C). See the minimum admission criterion below.

Because the MSDS promises to be an attractive option for students pursuing bachelor's degrees in certain areas, e.g. mathematics, computer science, actuarial science, statistics, physics, chemistry, and biology, and wish to obtain the bachelor's degree and a MS degree in data science in 5 years (U2G), we expect to also pursue approval for an eventual U2G option.

Mode and format: Face-to-face, 15-week classes per semester.

Location: Courses: on campus. Capstone project: can be off campus.

Minimum admission criteria: Bachelor's degree from a U.S. college or university, or an equivalent degree from another country. Minimum 3.0 GPA or the foreign equivalent. TOEFL/IELTS scores for international applicants without a U.S. degree and whose first language is not English. The TOEFL minimum is a score of 85 or higher on the Internet-based version of the test. The GRE is required as evidence of a student's quantitative aptitude.
Additional minimum admission requirements include:

- Two semesters of college calculus;

- One semester of college linear algebra;
- Prior formal coursework in computer science, equivalent to an introduction to computing course in the standard computer science curriculum.

Advising: Each enrolled student will be assigned an academic advisor to assist with the student's plan of study and to provide program mentoring.

Good-standing-criteria: We will follow the rules and regulations of the Graduate College; see https://www.grad.uiowa.edu/academics/rules-and-deadlines for detailed information.

Schedule and Program Length: Students can take classes on their own schedule. The intent is to schedule course offerings so that a student can finish the Master's degree in 1 year plus a semester (can be a summer semester). However, students may take courses at their own pace and can have up to 3 years to complete the degree requirements.

Hours required: 30 semester hours (10 courses).

Target Size of the Annual Cohort: The target for the initial cohort (in fall 2022) is between 10-20 students. The intent is to grow the program steadily from there, with a target incoming class of 40 each academic year.

Administration and Operation: The program will be administered and operated by the Department of Statistics and Actuarial Science with assistance provided by the College of Liberal Arts and Sciences (CLAS) staff.

The Department of Statistics and Actuarial Science will make admission decisions, develop the curriculum, provide student advising, and generally lead the academic aspects of the program. The existing college and departmental staff will provide logistical support for the program, including handling applications and enrollments, answering students' day-to-day questions, providing support at the classroom sites, and marketing the program.

To market and recruit for the program, the Department of Statistics and Actuarial Science will continue the efforts already in place for the MS in Statistics and the MS in Actuarial Science, which include online, email and social media outreach, and targeted traditional advertising. The Department of Statistics and Actuarial Science will particularly make effort to attract students from mathematics, computer science, actuarial science, physics, chemistry, biology, geoscience, etc. from Iowa and other universities in the Midwest. Because all these fields are becoming more and more data driven, there is a strong demand for data scientists with expertise in such fields.

4. Resource requirements

The program is designed to be self-sustaining in the Department of Statistics and Actuarial Science. The Department of Statistics and Actuarial Science will use and revise some of its current courses to satisfy the curriculum needs. Data science is a fast-developing field. Therefore, it is important to enhance the existing courses to reflect the advances in data science technologies, computing, statistical modeling and machine learning. This work can be done by the existing faculty in the Department of Statistics and Actuarial Science. However, if the enrollment in the program increases significantly over the target size of 40 in the future, the Department of Statistics and Actuarial Science will need to request additional faculty support from CLAS.

5. Implications

Implications for undergraduate work. As this program will only be available to graduate students, the program will have no direct implications for undergraduate work at the University. However, we anticipate indirect benefits for undergraduates in terms of shared course content.

Undergraduate-to-Graduate Option. We anticipate that the proposed MSDS program will be attractive to undergraduate students in a wide range of disciplines who are interested in pursuing MSDS in the U2G option. In particular, we will encourage undergraduate students in BS Data Science and related programs (e.g., students in mathematics, computer science and actuarial science) to participate in such a program. Interested students must follow the admissions guidelines and timetables for the existing U2G program of interest.

Implications for Other Programs at the University of Iowa.

We envision little to no adverse implications for existing programs at the University of Iowa, either at the undergraduate or graduate level, and we collected support letters (see Appendix A.4) from the University of Iowa Departments of Business Analytics, Biostatistics, and the Computer Science. A review of the ways in which existing programs compare to the proposed MSDS follows.

The most closely related—existing—master's degree is the Master of Science in Business Analytics (MSBA) program offered by the Tippie College of Business at the University of Iowa. It is a full-time business analytics program, offering courses in data visualization, databases, programming languages such as R and python from the business analytics perspective, and application areas such as text analytics, financial analytics, and healthcare analytics. This program trains students in advanced analytics techniques. It emphasizes developing skills to make students adept at using predictive

and prescriptive analytics to solve actual business problems. This program is sixteen months and includes an internship, capstone project class, and an emphasis on developing skills for clearly communicating results.

The UI Department of Computer Science (CS) offers a BA or BS and MS in Computer Science. In particular, the MS in CS has a machine learning and artificial intelligence bent to data science from the computer science perspective to interested students. Finally, the Interdisciplinary Graduate Program in Informatics (IGPI) is also housed in CS and has the aligned goals of providing students the background and skills to work/study at the interface of computing and specific discipline (for instance, bioinformatics, information science, geoinformatics, health informatics, and library science). The focus of IGPI is application oriented, emphasizing data acquisition, data manipulation, databases management and networking.

The proposed MSDS program is distinct from the aforementioned UI programs in the following aspects: (a) in contrast to domain-specific analytics programs, for instance, business analytics, health analytics and informatics, the proposed MSDS program is a generalist degree designed for students with a liberal arts BS degree and certain mathematical/analytical background (i.e., have taken and done well in two college calculus courses, one linear algebra course, one introductory computing course and have some basic coding skills in a computer programming language, such as C, Python and R). This target applicant pool is different from the applicant pool of MSBA, whose admission does not require college calculus and linear algebra; (b) the emphasis and the course design of the proposed MSDS program is to train students on statistical thinking and modeling aspects of data science, so that they not only learn the existing data science methods and technologies, but are also capable of learning and developing new technologies after graduation; (c) the employment opportunities of the graduates from the proposed MSDS program are positions in academe, industry, or government that involve experimental design, data visualization and wrangling, modeling, and analysis.

The proposed MSDS program, the Computer Science programs, the MS Business Analytics program and the IGPI program constitute complementary programs for training the entire spectrum of personnel needed for fully harnessing the data revolution to the benefits of the State of Iowa and the nation. Moreover, the education of the students from these diverse programs will be greatly enriched as the programs collaborate by allowing students to take elective courses across programs; see Appendix A.2.

The proposed MSDS program will also be beneficial to students in several other related programs such as Biostatistics, Educational Measurement and Statistics, and Geographical and Sustainability Sciences at the University of Iowa. The students in these programs are often involved in data-intensive work in their training and research. They can benefit from taking some of the courses in the MSDS program.

Related programs at other universities in the State of Iowa.

There are no existing MS in Data Science programs in the other Board of Regents Universities in Iowa. There are two undergraduate programs in data science: The College of Liberal Arts and Sciences at the Iowa State University offers a Bachelor of Science undergraduate degree in data science. The University of Northern Iowa has a B.A. Physics with Data Science Emphasis undergraduate program in the Department of Physics. The proposed MSDS is the only MS in data science program with an emphasis on statistical modeling and analysis in the State of Iowa.

6. External endorsement letters

We have reached out to Professor Ann Campbell, DEO, Department of Business Analytics; Professor Joseph Cavanaugh, DEO, Department of Biostatistics; and Professor Alberto Segre, DEO, Department of Computer Science, to seek their suggestions and comments for the proposed MSDS program, and which are incorporated into this proposal. We have also discussed with them future collaborations between the proposed MSDS program and the programs in business analysis, biostatistics and computer science in terms of sharing resources to benefit our related programs. Professor Campbell, Professor Cavanaugh and Professor Segre strongly support the proposed MSDS program.

Endorsement letters from them on behalf of the Department of Business Analytics, the Department of Biostatistics and the Department of Computer Science, University of Iowa are attached in Appendix A.4

7. List of Faculty

The following faculty in the Department of Statistics and Actuarial Science will be available to teach in the proposed program:

Matthew Bognar	Elias Shiu
Kung-Sik Chan	Nariankadu D Shyamalkumar
Rhonda DeCoook	Sanvesh Srivastava
Joyee Ghosh	Osnat Stramer
Jian Huang	Aixin Tan
Joseph B Lang	Luke Tierney
Ambrose Lo	Boxiang Wang
Lan Luo	Dale Zimmerman

8. Assessment of Future Needs, Commitments, and Opportunities

If this program proves successful, the Department of Statistics and Actuarial Science will consider expanding the enrollment of the program. A larger program would require more resources but also provide more opportunities to train the next generation of data scientists and serve the State of Iowa, the nation and the world.

9. Appendices

Appendix A.1: Curriculum Details for Core Courses

Master's First-Year Core Courses (6 courses totaling 19 semester hours)

Core learning objectives are to introduce students to: (1) foundations of probability and statistics; (2) data storage, access, and management; and (3) data visualization, exploration, modeling, analysis and uncertainty quantification.

Course Title	Existing?	Short Course Description
Probability and Statistics	Yes, STAT:3120	Foundations of probability and statistics
Applied Linear Regression	Yes, STAT:3200	Applied data modeling, analysis and uncertainty quantification
Statistical/Machine Learning	Yes, STAT: 4540	Unsupervised and supervised learning, prediction
Statistical Computing	Yes, STAT 5400	Data science computing with Python and R, data storage, access, and management
Data Visualization and Data Technologies	Yes, STAT:4580	Visualization and data technologies
Large Data Analysis	Yes, STAT: 4740	Big data analytics and machine learning

Probability and Statistics (STAT: 3120, 4 s.h.). Basic concepts of probability, statistical models, discrete and continuous random variables and their distributions, expectations, conditional expectations, estimation of parameters, testing statistical hypotheses.

Applied linear regression (STAT: 3200, 3 s.h.). Regression analysis with focus on applications; model formulation, checking, selection; interpretation and presentation of analysis results; simple and multiple linear regression; logistic regression; ANOVA; hands-on data analysis with computer software.

*Statistical Learning (STAT: 4540, 3 s.h.)***.** Introduction to supervised and unsupervised statistical learning, with a focus on regression, classification, and clustering; methods will be applied to real data using appropriate software; supervised learning topics include linear and nonlinear (e.g., logistic) regression, linear discriminant analysis, cross-validation, bootstrapping, model selection, and regularization methods (e.g., ridge and lasso); generalized additive and spline models, tree-based methods, random forests and boosting, and support-vector machines; unsupervised learning topics

include principal components and clustering. Requirements: an introductory statistics course and a regression course. Recommendations: prior exposure to programming and/or software, such as R, SAS, and Matlab.

Data Visualization and Data Technologies (STAT: 4580, 3 s.h.). Introduces common techniques for visualizing univariate and multivariate data, data summaries, and modeling results. Students will learn how to create and interpret these visualizations, and to assess effectiveness of different visualizations based on an understanding of human perception and statistical thinking. Data technologies for obtaining and preparing data for visualization and further analysis will also be discussed. Students will also learn how to present their results in written reports and to use version control to manage their work.

Computing in Statistics (STAT: 5400, 3 s.h.). Python, R; database management; graphical techniques; importing graphics into word-processing documents (e.g., LaTeX); creating reports in LaTeX; SAS; simulation methods (Monte Carlo studies, bootstrap, etc.).

Large data analysis (STAT: 4740, 3 s.h.). Current areas that deal with problem of Big Data; techniques from computer science, mathematics, statistics; high performance and parallel computing, matrix techniques, cluster analysis, visualization; variety of applications including Google PageRank, seismology, Netflix-type problems, weather forecasting; fusion of data with simulation; projects.

Master's Second-Year Core Courses (1 course totaling 2 semester hours)

Capstone Project. Each student will be supervised by a faculty member to complete a project that solves a real-world problem using knowledge gained from the core courses. Students are required to submit a written report and give an oral presentation of their projects; the written report must include the background and significance of the problem, analysis method, presentation and interpretation of the results including tables and visualization, discussion, and references, plus appendices comprising technical details and documentation of computer code used in the analysis. A capstone committee consisting of three faculty members will evaluate the capstone projects and assign the final grades (S or U), with inputs from the supervising faculty members.

Some example capstone projects:

(i) Fast Prediction of the Frequency of Tropical Cyclones

(ii) "All Things to All People": Modeling Attitudes to Social and Moral Issues in America and Western Europe Using Religious Attitudes

(iii) Understanding Pittsburgh Penguins' Offensive Strategies Using Spatial Analysis

(iv) Feedforward Neural Network for Predicting the Housing Market in Iowa City

Appendix A.2: Curriculum Details for Elective Offerings

Three electives from the following list (total 9 s.h.)

STAT:3210 Experimental Design and Analysis 3 s.h.
Single- and multifactor experiments; analysis of variance; multiple comparisons; contrasts; diagnostics; fixed, random, and mixed effects models; designs with blocking and/or nesting; two-level factorials and fractions thereof; use of statistical computing packages.

STAT:4520 Bayesian Statistics 3 s.h.
Bayesian statistical analysis, with focus on applications; Bayesian and frequentist methods compared; Bayesian model specification, choice of priors, computational methods; hands-on Bayesian data analysis using appropriate software; interpretation and presentation of analysis results.

STAT:4560 Statistics for Risk Modeling 3 s.h.
Theory and applications of general linear models, generalized linear models, and regression-based time series models; emphasis on parameter estimation, variable selection, and diagnostic checking for these models, and their use for statistical inference and prediction; practical implementations of these models to analyze actuarial and financial data.

STAT:5810 Research Data Management 3 s.h.
Introduction to data management techniques and problems encountered in gathering and processing data from biomedical investigations; introduction to SAS, techniques taught in SAS; designed for non-biostatistics majors. Offered fall and spring semesters. Recommendations: prior programming experience with C, C++, Python, Java, or other.

STAT:6530 Environmental and Spatial Statistics 3 s.h.
Geostatistics kriging, variogram estimation, trend estimation, sampling design, extensions to river networks and the globe, lattice data analysis, analysis of spatial point patterns.

STAT:6550 Introductory Longitudinal Data Analysis 3 s.h.
Introduction to statistical models and estimation methods for outcome variables (normal and non-normal) clustered or measured repeatedly in time or space; focus on applications and computer software methods for ANOVA based methods, hierarchical linear models, linear mixed models, correlated regression models, generalized estimating equations, and generalized linear mixed models. Offered fall semesters.

STAT:6560 Applied Time Series Analysis 3 s.h.

General stationary, nonstationary models, autocovariance autocorrelation functions; stationary, nonstationary autoregressive integrated moving average models; identification, estimation, forecasting in linear models; use of statistical computer packages. Offered spring semesters.

BIOS:6720 Statistical Machine Learning for Biomedical and Public Health Data 3 s.h.
Statistical machine learning techniques for analysis of biomedical and public health data; methodology and application of unsupervised learning, supervised learning for regression and classification, ensemble learning, model assessment, feature selection, and high-dimensional inference.

CS:4310 Design and Implementation of Algorithms 3 s.h.
Algorithm design techniques with emphasis on programming and implementation in the work of students; topics include data structures, graph algorithms, divide-and-conquer, dynamic programming, randomized algorithms, and dealing with intractability; primarily for master's degree students in computer science.

CS:5110 Introduction to Informatics 3 s.h.
Fundamentals of computer science: algorithms, complexity, relational databases, systems concepts, programming in Python.

CS:4470 Health Data Analytics 3 s.h.
Analysis of different kinds of health care data, such as patient electronic medical records, public health data, biomedical publications, social media pertaining to health, and ontologies in health care; students will read papers exploring different kinds of research and application development involving such data; course will run in distinct modules with each focused on a dataset type and related research; students must be comfortable with programming (e.g., Java, Python, Perl).

CS:4400 Database Systems 3 s.h.
Introduction to database systems including querying using SQL, design using ER diagrams, developing relational databases, programming web applications using PHP or JDBC.

CS:4440 Web Mining 3 s.h. Core methods underlying development of applications on the Web; examples of relevant applications, including those pertaining to information retrieval, summarization of Web documents, and identifying social networks.

CS:5430 Machine Learning 3 s.h.
Fundamental machine learning techniques as well as hands-on experience applying these techniques and developing new techniques for solving problems from the real world; topics include regression (least square regression, lasso), classification (naive

Bayes, nearest neighbor, support vector machines, logistic regression), kernel methods, unsupervised methods (k-means clustering, spectral clustering, dimensionality reduction), stochastic optimization, deep learning, and recent advances in big data analytics.

CS:5630 Cloud Computing Technology 3 s.h.
Explores infrastructure and programming paradigms of scalable systems and databases; provides experience with popular cluster frameworks (MapReduce, Hadoop, Spark, Flink, or similar) through programming exercises, projects, and experiments; assigned readings and case studies explore themes such as replication, data sharding, looser types of consistency, virtualization, consensus, and barrier synchronization; cloud system stacks developed by Google, Amazon, Facebook, and Microsoft.

BAIS:6100 Text Analytics 3 s.h.
Concepts and techniques of text mining; practice of using statistical tools to automatically extract meaning and patterns from collections of text documents; topics include document representation, text classification and clustering, sentiment analysis and topic modeling.

BAIS:6130 Applied Optimization 3 s.h.
Use of optimization (also called prescriptive analytics or mathematical programming) to make tactical and strategic decisions; advanced optimization skills including data collection and preparation, logical modeling, and solution interpretation and implementation within a software environment; applications in the various functional areas of business are discussed throughout.

BAIS:6210 Data Leadership and Management 3 s.h.
Core chief information officer (CIO) basics; focus on how to keep technology, systems, and procedures supporting business goal outcomes including management of information technology (IT)) teams, systems selection, vendor negotiation, change, information risk, data integrity, ethics, information system (IS) policies, strategies, cloud computing, and budget.

Appendix A.3: The Existing Data Science and Related Programs

In the Big Ten Academic Alliance (BTAA), the following institutions have data science programs: University of Iowa (BS), University of Michigan (BS and MS), University of Minnesota (MS), University of Wisconsin-Madison (BS statistics/data science option and M.S. statistics/data science option), Michigan State University (BS), Northwestern University (Online MSDS), Indiana University Bloomington (MS), Ohio State University (BS, MS in Data Analytics).

Among the BTAA institutions, six have business-oriented MA or MBA programs in business analytics (University of Iowa, Michigan State University, University of Minnesota, Northwestern University, Purdue University and Rutgers University), three have related informatics MA/MS programs (Indiana University-Bloomington, University of Michigan and Rutgers University), and five have one or more online masters-level programs of various types (University of Illinois Urbana-Champaign, Indiana UniversityBloomington, Pennsylvania State University, University of Wisconsin-Madison and Northwestern University; Indiana University-Bloomington and Pennsylvania State University each has two online programs).

Several institutions, including University of Wisconsin-Madison, University of Illinois Urbana-Champaign, Indiana University-Bloomington, University of Michigan, Pennsylvania State University, University of Minnesota, Rutgers University, University of Maryland, have both data science (MPS data science, MS data science and MS Statistics/Data Science options, residential) and business analytics (MS business or MBA with business analytics concentration, residential or online). Also, among BTAA institutions and nationally, a number of universities have both residential business analytics (under various names) and data science MS programs, including University of Minnesota (MSBA, MSDS), Carnegie Mellon University (Master of Information Systems Management, Business Intelligence and Data Analytics Concentration; Masters of Business Administration, Business Analytics Track; Masters in Computational Data Science; Masters in Machine Learning), Columbia University (Master of Science in Applied Analytics, MSDS), Duke University (Master of Quantitative Management, MSDS), University of Virginia (Master of Science in Commerce: Business Analytics Track, MSDS), University of Washington (MSBA, MSDS).

Nationally, many schools offer MSDS programs; the Institute for Advanced Analytics lists nearly 250 programs offering masters degrees in Analytics or Data Science at universities in the US, producing about 8,000 to 10,000 graduates annually[12]. The website master in data science[13] identifies the following 23 schools as having the best master's data science programs in the United States: University of California-Berkeley;

[12] https://analytics.ncsu.edu/?page_id=4184

[13] https://www.mastersindatascience.org

Drexel University, Illinois Institute of Technology, University of Missouri-Columbia, Columbia University, Indiana University-Bloomington, Johns Hopkins University, Northeastern University, University of California-San Diego, University of CaliforniaSanta Barbara, University of Michigan, University of Virginia, University of WisconsinMadison, University of Denver, Syracuse University, Southern Methodist University, Bay Path University, Cabrini University, Carleton College, Carnegie Mellon University, Cornell University, Duke University, Harvard University.

Appendix A.4: Endorsement Letters

Tippie College of Business
Department of Business Analytics

University of Iowa
108 John Pappajohn Business Building, W252
Iowa City, Iowa 52242-1994
319-335-0858
Fax 319-335-
1956
tippie.uiowa.edu

February 15, 2021

Dear Kung Sik:

I am writing this letter as DEO of the Department of Business Analytics in the Tippie College of Business.

I support the proposal for the MS in Data Science to be offered by the Department of Statistics and Actuarial Science.

The Department of Business Analytics offers a related master's degree in our on-campus program: the Masters of Science in Business Analytics. Our degree helps students learn how to use predictive and prescriptive analytics to solve business problems. Our degree involves courses in data visualization, databases, programming languages such as R and python, and application areas such as text analytics, financial analytics, and healthcare analytics. There is definite overlap in the topics of some of these courses with the courses in the Data Science program. I had a great conversation with you and Jian Huang about how to make the differences in these classes and the programs clear for our students to prevent confusion.

There are many benefits from having multiple data-focused programs on campus. All programs benefit from more electives. We intend to share appropriate electives, so students from both programs can have more choices.

More programs with a data-focus can also improve the University of Iowa's reputation as a destination for data science research. This can improve the demand from students and help with recruiting excellent faculty in this area. This can also lead to more cross-collaborations between students and faculty in both of our departments.

Regards,

Ann Melissa Campbell (she/her/hers)
DEO, Department of Business Analytics
Tippie Research Professor
W220 Pappajohn Business Building, Iowa City, Iowa 52242
Office: 319-335-0918 **ann-campbell@uiowa.edu**

THE UNIVERSITY OF IOWA
College of Public Health
Department of Biostatistics

University of Iowa
145 N Riverside Drive
100 College of Public Health, N300
Iowa City, Iowa 52242 319-384-1582 https://www.public-health.uiowa.edu/

February 14, 2021

Kung-Sik Chan, PhD
Robert V. Hogg Professor and Chair
Department of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall

Dear Kung-Sik,

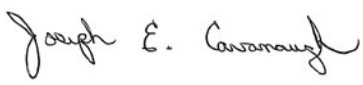
I am writing to express my strong support for the proposal that has been prepared by the Department of Statistics and Actuarial Science for an MS in Data Science.

As DEO of Biostatistics, I can attest to the rapid growth and evolution of Data Science, a discipline that is integrated and aligned with the statistical sciences, and also borrows heavily from the computational sciences. The University of Iowa would greatly benefit from offering an MS degree in this area. The Department of Statistics and Actuarial Science is the natural home for such a degree for several reasons. First and foremost, their faculty have the appropriate expertise in modern statistical and computational methods to offer such a degree. Second, as outlined in the proposal, the Department already offers courses that can serve as the core for the degree. Third, the Department has already developed a successful undergraduate degree in Data Science in partnership with the Department of Computer Science.

The Department of Biostatistics has a close, symbiotic relationship with Statistics and Actuarial Science, and the union of courses offered by both departments greatly enhances graduate degrees in both Statistics and Biostatistics. Our faculty will make every effort to assist our colleagues as they develop this MS program to ensure that it is successful.

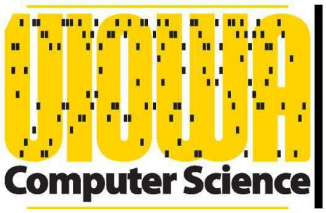
I am hopeful that this degree proposal is approved, since it will address an existing need among the University's degree programs, and allow the University to remain on the forefront of the statistical sciences.

Regards,



Joseph E. Cavanaugh, PhD

Professor and Head of Biostatistics
Professor of Statistics and Actuarial Science (Secondary)



COLLEGE OF
LIBERAL ARTS & SCIENCES
Department of Computer Science
14 MacLean Hall
Iowa City, Iowa 52242-1419
319-335-0713 *Fax* 319-335-3624
www.cs.uiowa.edu

Kung-Sik Chan, PhD
Robert V. Hogg Professor and Chair
Department of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall

February 18, 2021

Dear Kung-Sik;

I'm writing to express my support for your proposed Master of Science degree in Data Science.

Alberto Maria Segre alberto-segre@uiowa.edu 319/335-1713 Tel

The Department of Computer Science offers several related programs. In addition to the PhD, our Department has a thriving Master of Computer Science (MCS) degree program, where students can learn to be effective software developers, including algorithmic concepts, machine learning, optimization, simulation, programming language design, formal methods, human interface design, and a wide array of more application-oriented topics. More recently, the Department of Computer Science has also become the home of record for the Interdisciplinary Informatics Graduate Program (IGPI), which includes MA/MS degrees in a broad range of domain-oriented areas, including, *e.g.*, geoinformatics, health informatics, and so on. It is therefore not surprising that several of our existing graduate courses appear in your proposal! Moreover, our undergraduate programs (the BA/BS in Computer Science, BA/BS in Informatics, the BSE in Computer Science and Engineering, and the BS in Data Science) all present attractive feeder opportunities for your new proposed program.

I applaud your Department's proposal to offer a data analytics program that is grounded in sound statistical methods, and I am fully supportive of this initiative.

Sincerely yours,

Alberto Maria Segre
Professor and Chair
Gerard P. Weeg Faculty Scholar in Informatics