Dr. Gordon J. Miller
Chair, LAS Curriculum Committee

On behalf of the Data Science Curriculum committee, we are submitting a program proposal for an undergraduate minor in Data Science for your consideration. The Data Science Curriculum committee consists of the following members from across ISU colleges.

1. David G. Acker, Associate Dean, Global Agriculture Programs
2. Raymond W Arritt, Dept. of Agronomy
3. Cameron Campbell, Associate Dean, College of Design
4. Bryan G Clendenen, Dept. of Graphic Design
5. Carolyn J Dill, Dept. of Genetics/Development & Cell Biology
6. Cassandra J Dorius, Dept. of Human Development & Family Studies
7. Shawn Dorius, Dept. of Sociology
8. Arne Hallam, Associate Dean, College of Liberal Arts and Sciences
9. Chinmay Hegde, Dept. of Electrical and Computer Engineering
10. Heike Hofmann, Dept. of Statistics
11. David K Holger, Associate Provost and Dean of Graduate College
12. Matthew Hufford, Dept. of Ecology, Evolution & Organismal Biology
13. Danny J Johnson, Associate Dean, College of Business
14. Wolfgang H Kliemann, Associate Vice President for Research
15. Sigurdur Olafsson, Dept. of Industrial & Manufacturing Systems Engineering
16. Francis Y Owusu, Dept. Chair, Dept. of Community and Regional Planning
17. James M Reecy, Department of Animal Science
18. Sarah M Ryan, Dept. of Industrial & Manufacturing Systems Engineering
19. Sree Nilakanta, Dept. of Supply Chain & Information Systems
20. Hridesh Rajan, Dept. of Computer Science
21. Heather L Rouse, Dept. of Human Development & Family Studies
22. Sriram Sundararajan, Associate Dean, College of Engineering
23. Wallapak Tavanapong, Dept. of Computer Science

The data science curriculum committee formed a subcommittee and charged it with designing and developing a program proposal for a data science minor. This subcommittee consisted of Raymond Arrit, Arne Hallam, Chinmay Hegde, Kevin Kane, Dengpan Liu, Hridesh Rajan and Sriram Sundararajan. The attached proposal is the work of this subcommittee with regular feedback and approval from the data science curriculum committee.

Respectfully submitted,
Hridesh Rajan, Arne Hallam, and Sriram Sundararajan

Program Proposal for a Minor

1. **Name of the proposed minor.**

   Data Science

2. **Name of the academic units involved.**

   College of Liberal Arts and Sciences

   College of Engineering

   College of Agriculture and Life Sciences

   College of Business

   College of Design

   College of Human Sciences

3. **Name of contact person(s).**

   Dr. Hridesh Rajan, Department of Computer Science (hridesh@iastate.edu). Dr. Rajan is the chair of the Data Science Minor and Certificate subcommittee of the ISU data science curriculum committee.

   Dr. Arne Hallam, Associate Dean, College of Liberal Arts and Sciences (ahallam@iastate.edu). Dr. Hallam is the co-chair of the ISU data science curriculum committee.

   Dr. Sriram Sundararajan, Associate Dean, College of Engineering (srirams@iastate.edu). Dr. Sundararajan is the co-chair of the ISU data science curriculum committee.

4. **General description of the undergraduate minor.**

   The proliferation of rich and complex data in science, industry and government is fueling the rapid growth of data science as a discipline. The National Science Foundation Directorate for Mathematical and Physical Sciences Support for the Statistical Sciences at NSF—a subcommittee of the Mathematical and Physical Sciences Advisory Committee—states that data science is the "*science of planning for, acquisition, management, analysis of, and inference from data*"[1]. The minor is intended for students studying any discipline at Iowa State University with the goal of enabling them to work in data science. The minor consists of nine credits hours of data science core courses and six credit hours in data science electives. The courses in the data science minor are designed to provide students with the requisite background that would enable them to take jobs with significant data science components, e.g., establishing and operating data analysis pipelines[2].

[1] Iain Johnstone and Fred Robert (chairs), "Data Science at NSF", April 2014. https://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf

[2] A data analysis pipeline refers to a conceptual framework, often fully or semi-automated using computational tools, for solving data science problems that consumes data from one end and produces insights from the other end.

5. Need for the proposed minor.

There is a massive demand for data science professionals today both in Iowa and nationally, and this market is expected to grow. Harvard Business Review labels data scientist as the most attractive job of the 21st century[3]. A report from the McKinsey Global Institute forecasts the need for hundreds of thousands of data science jobs in the next decade[4]. Iowa State University is well positioned to fill the need due to its long history in data-driven sciences. President Leath's significant recent investments in the area have served to strengthen ISU's capacity in this field further. The minor is designed to complement existing undergraduate degrees and prepare students who have an in-depth understanding of data science in a variety of domains. Understanding of the domain and context is considered a critical skill for a data scientist and this minor's design recognizes this.

6. Objectives of the proposed undergraduate minor including the student learning outcomes and how the learning outcomes will be assessed.

Objectives

The minor in data science will prepare students with the technical and communication skills to enter the workforce as domain *experts* with data science *skills.* These positions are in high demand today in industry, nonprofit agencies, and government.

Learning Outcomes

After completing the minor in data science, students will demonstrate

a. an ability to apply data science concepts, tools and methods to data analysis pipelines,

b. an understanding of ethical, legal, societal, and economic concerns in application of data science concepts,

c. an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders, and

d. an ability to function on multi-disciplinary teams using concepts and tools from data science.

Assessment

To assess the learning outcomes for data science minor, the proposed minor will use DS 201, DS 202 and DS 301 as shown in the table below. These are the three core courses in the data science minor. Experimental course proposal forms for these courses are attached to this minor proposal. These course proposals will be submitted in Spring 2017 for approval.

Outcome	Courses in which the outcome will be assessed
a.	DS 202: data acquisition and exploratory data analysis

[3] https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[4] McKinsey Global Institute, "*Big Data: The Next Frontier for Innovation, Competition, and Productivity*", 2011 New York: McKinsey & Co. http://www.mckinsey.com/business-functions/digital-mckinsey/our- insights/big-data-the-next-frontier-for-innovation

b.	DS 201: introduction to data science DS 301: applied data modeling and predictive analysis
c.	DS 202: data acquisition and exploratory data analysis
d.	DS 301: applied data modeling and predictive analysis

7. Relationship of the minor to other programs at Iowa State University.

There are no other undergraduate programs at ISU that meet the objectives of the data science minor. Majors, minors and/or individual courses in ISU colleges, e.g. computer science, statistics, business analytics, bioinformatics, etc., provide material relevant and useful for those pursuing careers in data science but none provide the breadth of coverage open to a wide range of students as does this minor. Some topics covered in the three core courses in the minor are not offered at ISU, e.g. principal components of a data science pipeline, data science project management, ethical issues in data science. Some topics covered in this minor are covered only in specialized courses available only to individuals within a very small number of majors due to prerequisite requirements, e.g. elements of predictive analysis such as training and test sets; feature extraction; survey of algorithmic machine learning techniques. A proposal for a data science certificate is concurrently being routed through for evaluation.

8. Relationship of the undergraduate minor to the strategic plans of the university, of the college, and of department or program.

Data science is a priority area for Iowa State University. For example, ISU President's Destination 2050 initiative features Big Data as one of the six targets[5]. Iowa State University's Presidential Initiative for Interdisciplinary Research has specifically targeted data driven science[6]. Data science is also one of the strategic areas in several ISU colleges and academic units including LAS[7].

9. Comparison of the proposed undergraduate minor with similar programs at other universities, including the Regent's universities.

As of this writing, there are no such programs at the Regent's universities. The University of Iowa has a graduate program in business analytics whereas the current proposal is for undergraduate students. The University of Iowa has also introduced a track within their Statistics B.S. degree called "Statistical Computing and Data Science," whereas the current proposal is for an undergraduate minor. The University of Iowa has also introduced a certificate called "Large Data Analysis Certificate" starting Fall 2015. This certificate includes such existing courses as database management, high performance computing, optimization techniques, and knowledge discovery, whereas the proposed minor incorporates courses that are designed for data science and focuses on topics such as data acquisition and preparation, exploratory data

[5] http://www.destination2050.iastate.edu

[6] http://www.vpresearch.iastate.edu/index.cfm/47530/40579/presidential_research_initiative_promotes_big_thinking_in_datadriven_science

[7] https://news.las.iastate.edu/2017/01/11/college-of-liberal-arts-and-sciences-designated-as-recipient-of-one-of-iowa-state-universitys-largest-gifts/

analysis, visualization, predictive analysis, machine learning, etc., that more closely relate to setting up and operating data analysis pipelines. The University of Northern Iowa doesn't currently have any related programs.

10. Program requirements and procedures, including:

Organization of the minor: The table below shows the organization of the program. The overall program is divided into two categories of courses: Core Courses (9 credits), and Elective Courses (6 credits). At the time of this writing, the list of electives has a few representative courses, but as the program matures we anticipate including additional choices for the elective courses, in particular, some courses that have few prerequisites other than the core data science courses. For a course to qualify as an elective for the program, it must cover one or more of the learning outcomes of the program. The data science curriculum committee (see governance document) will approve new elective courses, and review them periodically to ensure that they continue to meet the learning outcomes.

Course	Credits	Title	Status
Core Courses (9 credits)			
DS 201	3	Introduction to Data Science	New
DS 202	3	Data Acquisition and Exploratory Data Analysis	New
DS 301	3	Applied Data Modeling and Predictive Analysis	New
Elective (6 credits) [more courses to be added later*]			
CRP 251X	3	Introduction to Geographic Information Systems	Existing
STAT 301	3	Intermediate Statistical Concepts and Methods	Existing
COMS 311	3	Design and Analysis of Algorithms	Existing
IE 312	3	Optimization	Existing
ABE 316	3	Applied Numerical Methods for Agricultural and Biosystems Engineering	Existing
BCBIO 322	3	Introduction to Bioinformatics and Computational Biology	Existing
STAT 330	3	Probability and Statistics for Computer Science	Existing
ENGL 332	3	Visual Communication of Quantitative Information	Existing
CRP 351X	3	Intermediate Geographic Information Systems	Existing
COMS 363	3	Introduction to Database Management Systems	Existing
MKT 368X	3	Marketing Analytics	Existing
ECON 371	3	Introductory Econometrics	Existing
STAT 407	3	Methods of Multivariate Analysis	Existing
LING 410X	3	Language as Data	Existing
STAT 430	3	Empirical Methods for the Computational Sciences	Existing
CprE 419	3	Software Tools for Large Data Analytics	Existing
COMS 424	3	Introduction to High Performance Computing	Existing
CPRE 426	3	Introduction to Parallel Algorithms and Programming	Existing
COMS 435	3	Algorithms for Large Data Sets: Theory and Practice	Existing
MIS 436	3	Introduction to Business Analytics	Existing
MIS 446	3	Advanced Business Analytics	Existing

FIN 450X	3	Analytical Finance	Existing
CRP 452	3	Geographic Data Management and Planning Analysis	Existing
COMS 453X	3	Privacy Preserving Algorithms and Data Security	Existing
CRP 456	3	GIS Programming and Automation	Existing
STAT 457	3	Applied Categorical Data Analysis	Existing
COMS 474	3	Introduction to Machine Learning	Existing
STAT 480	3	Statistical Computing Applications	Existing
IE 483	3	Knowledge Discovery and Data Mining	Existing

*All electives will address one or more of the outcomes a, b, and c.

a. prerequisites for prospective students;

The minor is intended for students pursuing a baccalaureate degree at ISU who have taken at least one year each of high school algebra, geometry, and advanced algebra. This requirement is consistent with ISU admissions requirement.

In addition, DS 301 has a prerequisite of having taken at least one university level statistics course and DS 201. DS 202 has a prerequisite of having taken DS 201.

b. application and selection process;

As per ISU rules, the minor is available only to students pursuing a baccalaureate degree at ISU. A student may declare for the minor according to normal university policies[8].

c. language requirements;

No additional language requirements

d. courses and seminars presently available for credit toward the data science minor;

CRP 251X: Introduction to Geographic Information Systems

STAT 301: Intermediate Statistical Concepts and Methods

COMS 311: Design and Analysis of Algorithms

IE 312: Optimization

ABE 316: Applied Numerical Methods for A B E

BCBIO 322: Introduction to Bioinformatics and Computational Biology

[8] See "Degree Planning" in the catalog at http://catalog.iastate.edu/academics/#degreeplanningtext

STAT 330: Probability and Statistics for Computer Science

ENGL 332: Visual Communication of Quantitative Information

CRP 351X: Intermediate Geographic Information Systems

COMS 363: Introduction to Database Management Systems

MKT 368X Marketing Analytics

ECON 371: Introductory Econometrics

STAT 407: Methods of Multivariate Analysis

LING 410X: Language as Data

STAT 430: Empirical Methods for the Computational Sciences

CprE 419: Software Tools for Large Data Analytics

COMS 424: Introduction to High Performance Computing

CPRE 426: Introduction to Parallel Algorithms and Programming

COMS 435: Algorithms for Large Data Sets: Theory and Practice

MIS 436: Introduction to Business Analytics

MIS 446: Advanced Business Analytics

FIN 450X Analytical Finance

CRP 452: Geographic Data Management and Planning Analysis

COMS 453X: Privacy Preserving Algorithms and Data Security

CRP 456: GIS Programming and Automation

STAT 457: Applied Categorical Data Analysis

COMS 474: Introduction to Machine Learning

STAT 480: Statistical Computing Applications

IE 483: Knowledge Discovery and Data Mining

e. proposed new courses or modifications of existing courses;

The following three new courses will be added. DS 201 will be first offered in Fall 2017 or after approval of the program, whichever is later. DS 202 and DS 301 will be first offered in Spring 2018 or after approval of the program, whichever is later.

DS 201: Introduction to Data Science. Cr. 2-2. F.S.

Prereq: 1-1/2 years of high school algebra

Data Science concepts and their applications; domain case studies; overview of data analysis; major components of data analysis pipelines; computing concepts for data science; descriptive data analysis; hands-on data analysis experience; communicating findings to stakeholders, and ethical issues in data science.

DS 202: Data Acquisition and Exploratory Data Analysis. Cr. 2-2. F.S.

Prereq: DS 201

Data acquisition: file structures, web-scraping, database access; ethical aspects of data acquisition; types of data displays; numerical and visual summaries of data; pipelines for data analysis: filtering, transformation, aggregation, visualization and (simple) modeling; good practices of displaying data; data exploration cycle; graphics as tools of data exploration; strategies and techniques for data visualizations; basics of reproducibility and repeatability; web-based interactive applets for visual presentation of data and results.

DS 301: Applied Data Modeling and Predictive Analysis. Cr. 3-0. F.S.
Prereq: DS 201, and one of STAT 101, 104, 105, 201, 226, 231, 305, 322, 330
Elements of predictive analysis such as training and test sets; feature extraction; survey of algorithmic machine learning techniques, e.g. decision trees, Naïve Bayes, and random forests; survey of data modeling techniques, e.g. linear model and regression analysis; assessment and diagnostics: overfitting, error rates, residual analysis, model assumptions checking; communicating findings to stakeholders in written, oral, verbal and electronic form, and ethical issues in data science.

f. advising of students;

Major specific advising will be handled by existing advisers in primary majors. Minor specific advising as a supplement to the responsibilities of primary major advisers will be handled by the data science specific advisers provided by the college of Liberal Arts and Sciences.

g. implications for related areas within the university.

- The proposed program will enhance the demand for currently offered courses. In particular, the demand for all of the data science electives will increase.
- The proposed program will provide additional documentation of knowledge for students in a large number of current majors, some closely related to data science such as Statistics, Computer Science, Industrial and Manufacturing Engineering, Computer Engineering, Management Information Systems. It will also open avenues for learning important skills in data science for students in other disciplines such as English, Journalism, History, Design, Marketing, Education, Biology, and so forth.

11. General description of the resources currently available and future resource needs, in terms of:

a. faculty members;

Faculty positions for teaching the existing elective courses listed in 10.d will continue to be needed to sustain the data science minor. As enrollment expands, new resources or reallocation of available resources will be necessary.

The College of Liberal Arts & Sciences will commit resources from within the college that will be needed to propose, develop, and teach the three new data science courses listed in 10.e (see attached support letter from the LAS Dean).

The revenue generated from teaching the data science courses are anticipated to offset the expended resources. These resources may come from new hires or reallocation of existing funds within the relevant units.

To the extent that faculty members outside of LAS are involved in teaching the four core courses, other colleges will need to provide the resources to support those faculty members.

b. computers, laboratories, and other facilities;

Existing facilities associated with existing courses e.g. data science electives will continue to be used. As the program expands there may be a need for new computer/statistics lab facilities. CAC funds will cover additional expenses associated with computers and undergraduate personnel.

c. research facilities

No new research facilities are needed.

d. library facilities (journals, documents, etc.) in the proposed area;

No new resources needed.

e. supplies, field work, student recruitment, etc.

While this program doesn't have significant supply needs at this time, a small supply budget will be required for various needs, e.g. printing information brochures to be used at college orientation events to recruit for data science introductory courses. LAS will provide this fund.

12. Describe the needs for new resources and/or reallocated resources. Attach to the program proposal memos from the department chair(s), the college dean(s), and

other appropriate persons, agreeing to the allocation of new resources and/or the reallocation of resources.

This minor will require two reallocations of resources. First, reallocation of resources would be required to propose, develop, and teach four new courses listed in 10.e. It is anticipated that the revenue generated from teaching the four new courses would offset this reallocation in the long run. Second, time of the curriculum committee will be reallocated toward the data science minor program. There will also need to be some resources allocated to providing minor specific advising as a supplement to the responsibilities of primary major advisers. LAS will provide the initial resources for this purpose.

13. Attach to the program proposal, letters of support, recommendations, and statements when appropriate, from programs and departments at ISU which are associated with the proposed program or have an interest in the proposed program.

Letters of support are being collected from the Deans of involved colleges.

14. If the new program is interdisciplinary, a governance document should be created and submitted to the Associate Provost for Academic Programs. Indicate here that it has been completed.

A governance document has been created and is being discussed with the Associate Provost for Academic programs.

New Experimental Course Proposal

DS 201: Introduction to Data Science. Cr. 2-2. F.S.

Instruction Type (Hours per Week): Lecture 2; Lab 2

Prerequisites: *1-½ years of high school algebra*

Description:
Data Science concepts and their applications; domain case studies with applications in various fields; overview of data analysis; major components of data analysis pipelines; computing concepts for data science; descriptive data analysis; hands-on data analysis experience; communicating findings to stakeholders, and ethical issues in data science.

Reason for proposal:

The proliferation of rich and complex data in science, industry and government is fueling the rapid growth of data science as a discipline. The National Science Foundation Directorate for Mathematical and Physical Sciences Support for the Statistical Sciences at NSF—a subcommittee of the Mathematical and Physical Sciences Advisory Committee—states that data science is the "*science of planning for, acquisition, management, analysis of, and inference from data*"[9]. This introductory course in data science is built as the first of three core courses for the Data Science Undergraduate Minor. The main purpose of this course is to allow undergraduate level students to capture and work with data. Students will be able to understand Data Science concepts and their use; work with Data Science case studies; understand the basic skills and principles of Data Analysis; identify the major components of data analysis pipelines; apply programming languages for problems in data science; perform descriptive data analysis; participate in hands-on data analysis experiences; communicate findings to stakeholders; and understand ethical issues in data science.

Course outcomes/objective:

After completing this course, students should be able to:

1. Define key ideas in data science and data analysis
2. Identify major components and various phases of a data science pipeline
3. Construct programs in a high level language to solve basic data science problems
4. Identify major components of descriptive data analysis
5. Define key ideas in data science project management and identify ethical issues in data science

Course content/major topics to be addressed:

Major topics include:

- Introduction – What is Data Science? Motivating case studies.
- Overview of Data Analysis

[9] Iain Johnstone and Fred Robert (chairs), "Data Science at NSF", April 2014. https://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf

    - Major components of a data analysis pipelines
    - Exploration of the data science process
- Computing for data science
    - data types
    - operations
    - control structures
    - functions
    - scoping rules
    - iterations
    - running and debugging programs.
- Descriptive Data Analysis
    - The fundamental concepts and methods of statistics
    - How to interpret graphical and numerical summaries of data
    - Reason and assumptions behind the calculations
    - The correct interpretation of results
- Hands-on data analysis experience
- Communication findings to stakeholders
    - Written, oral, verbal and electronic forms of communication
- Ethical issues in data science

Assessment Plans:

Student performance will be measured via programming assignments, laboratory exercises, and written and programming exams.

Relationship of this course to existing courses in other departments and programs (supporting, overlap, etc.):

There are no existing courses that are teaching introductory data science at Iowa State University. There are, however, few courses that cover aspects of data science in computer science and statistics courses, and analytics in IMSE, and business. These are either upper-level undergraduate courses or graduate courses with several prerequisites.

The DS 201 is a core course for the Undergraduate Data Science Minor which cover a broader range of Data Science areas and serve as the first core course covering the construction and use of Data Science pipelines.

New Experimental Course Proposal

DS 202: Data Acquisition and Exploratory Data Analysis. Cr. 2-2. F.S.

Instruction Type (Hours per Week): Lecture 2; Lab 2

Prerequisites: *DS 201*

Description:
Data acquisition: file structures, web-scraping, database access; ethical aspects of data acquisition; types of data displays; numerical and visual summaries of data; pipelines for data analysis: filtering, transformation, aggregation, visualization and (simple) modeling; good practices of displaying data; data exploration cycle; graphics as tools of data exploration; strategies and techniques for data visualizations; basics of reproducibility and repeatability; web-based interactive applets for visual presentation of data and results. Programming exercises.

Reason for proposal:

This introductory course in data science is built as the second of three core courses for the Data Science Undergraduate Minor. This course is designed to emphasize the importance of data acquisition and exploratory data analysis in a data science lifecycle. Much of the efforts in practical data science projects are spent in data acquisition and understanding the data via exploratory data analysis and this course will emphasize those concepts. Students will be able to understand various data formats, setup pipelines for data analysis that filter, transform, aggregate and visualize data. Students will also be able to perform (simple) modeling of data. Students will also learn about good practices of displaying data, data exploration cycle. Students will also learn about the basics of reproducibility and repeatability. Students will participate in hands-on data analysis experiences; communicate findings to stakeholders; and understand ethical issues in data science.

Course outcomes/objective:

After completing this course, students should be able to:

1. Define key file storage structures and methods to acquire data from those structures
2. Identify different types of data displays and the purpose for the same
3. Construct a data analysis pipeline in a high-level language to solve basic data science problem involving steps such as filtering, transformation, aggregation, visualization
4. Apply strategies and techniques for data visualizations
5. Define key ideas in reproducibility and repeatability and identify ethical issues in data science

Course content/major topics to be addressed:

Major topics include:

- Introduction – What is Data Acquisition? What is Exploratory Data Analysis? Motivating case studies.
- Data acquisition

    - file structures
    - web-scraping
    - database access
    - ethical aspects of data acquisition
- Types of data displays
- Numerical and visual summaries of data
- Pipelines for data analysis
    - filtering
    - transformation
    - aggregation
    - visualization and (simple) modeling
- Best practices for displaying data
- Data exploration cycle
- Graphics as tools of data exploration
- Strategies and techniques for data visualizations
- Hands-on data analysis experience
- Communicate findings to stakeholders
    - Written, oral, verbal and electronic forms of communication
- Basics of reproducibility and repeatability

Assessment Plans:

Student performance will be measured via programming assignments, exercises, and written and programming exams.

Relationship of this course to existing courses in other departments and programs (supporting, overlap, etc.):

There are no existing courses that are teaching introductory data acquisition and exploratory data analysis concepts at Iowa State University. There are, however, a few courses that cover aspects of exploratory data analysis. These are either upper-level undergraduate courses or graduate courses with several prerequisites.

New Experimental Course Proposal

DS 301: Applied Data Modeling and Predictive Analysis. Cr. 3-0. F.S.

Instruction Type (Hours per Week): Lecture 3; Lab 0

Prerequisites: *DS 201, one of STAT 101, 104, 105, 201, 226, 231, 305, 322, 330*

Description:
Elements of predictive analysis such as training and test sets; feature extraction; survey of algorithmic machine learning techniques, e.g. decision trees, Naïve Bayes, and random forests; survey of data modeling techniques, e.g. linear model and regression analysis; assessment and diagnostics: overfitting, error rates, residual analysis, model assumptions checking; communicating findings to stakeholders in written, oral, verbal and electronic form, and ethical issues in data science. Participation in a multi-disciplinary team project.

Reason for proposal:

This upper-level course in data science is built as the third of three core courses for the Data Science Undergraduate Minor. There is a large demand for data science professionals today both in Iowa and nationally. Machine learning and statistics provide the technical basis of data science. The course aims to provide an introduction to the basic tools and techniques of machine learning and statistics used in practical data analysis. The course is intended for students studying any discipline at Iowa State University with the goal of enabling them to apply machine learning and statistical techniques to analyze and understand their data.

Course outcomes/objective:

Upon successful completion of the course, students will have an understanding of the basic elements of predictive analysis and learn a variety of machine learning techniques such as linear models, decision trees, Naïve Bayes, and random forests. Students should be able to:

1. Define key ideas in predictive analysis such as training and test sets, and feature extraction
2. Apply basic machine learning techniques to solve data science problems
3. Evaluate and diagnose learned data models
4. Understand ethical concerns in data science
5. Communicate the output of data analysis pipelines to stakeholders
6. Participation in a multi-disciplinary teams

Course content/major topics to be addressed:

Students will learn to apply predictive analysis techniques. They will write programs to solve problems in hands-on Data science projects using machine learning and data modeling techniques. Major topics include:

- Introduction – What is predictive analysis? Motivating case studies.
- Elements of predictive analysis
    - training and test sets
    - feature extraction.
- Machine learning and data modeling techniques
    - linear model
    - decision trees
    - Naïve Bayes
    - random forests
    - regression analysis.
- Model assessment and diagnostics
    - overfitting
    - error rates
    - residual analysis
    - model assumptions checking.
- Selected applications in various domains
- Ethical issues in data science

Assessment Plans:

Student performance will be measured via programming assignments, exercises, and written and programming exams.

Relationship of this course to existing courses in other departments and programs (supporting, overlap, etc.):

There are no existing courses that are teaching predictive analysis and data modeling in an applied manner at the undergraduate-level at Iowa State University. This course may have small overlap on survey of the basic machine learning techniques with COM S 474 Introduction to Machine Learning and IE 483 Knowledge Discovery and Data Mining. COM S 474 focuses on machine learning techniques and mainly targets students in computer science and related disciplines. IE 483 focuses on algorithm techniques that can be used for data mining tasks in manufacturing and service industries. This course targets students from any disciplines interested in Data Science. Students could further their knowledge in machine learning and data mining by taking COMS 474 or IE 483 after taking this course.

Letter from the Computer Curriculum Coordination Committee

From: Hallam, Arne [LAS]
Sent: Thursday, February 2, 2017 2:47 PM
To: Rajan, Hridesh [COM S]; Miller, Gordon J [CHEM]
Cc: Hallam, Arne [LAS]
Subject: Data Science Certificate and Minor

The Computer Curriculum Coordination Committee discussed the Data Science Certificate and Minor on 5 December 2016.
The requirements for the minor and certificate were discussed.
The committee also discussed Data Science 201, 202, 301 and 401.
The committee has no concerns with proposals.
You can proceed with your blessing.

_____

Arne Hallam, Professor of Economics and Associate Dean
College of Liberal Arts and Sciences
237 Catt Hall
Iowa State University
Ames, IA 50011

ahallam@iastate.edu
Work: 515-294-5861
FAX: 515-294-1303
Home: 515-292-8739

Academic Program Approval Voting Record

This document is to be appended as the last page of the proposal for any new or revised academic program to record the successive votes of approval as the proposal moves through its required review and approval steps. Consult Faculty Handbook Section 10.8 or the Faculty Senate Curriculum Committee website for information regarding Committee review and voting requirements for each action.

Curricular Action: (check appropriate boxes below)

1. X New Program □ Name Change □ Discontinuation □ Concurrent Degree for:
2. □ Undergraduate Major □ Graduate Major X Undergraduate Minor □ Graduate Minor
   □ Undergraduate Certificate □ Graduate Certificate □ Other: _____
3. Name of Proposed Change: Data Science
4. Name of Contact Person: Hridesh Rajan, Arne Hallam, and Sriram Sundararajan

   e-mail address: hridesh@iastate.edu, ahallam@iastate.edu, and srirams@iastate.edu
5. Primary College: LAS Secondary College: _____
6. Involved Department(s): _____ _____

   _____ _____

Voting record for this curricular action:

Voting Body	Votes			Date of Vote
	For	Against	Abstain	
Data Science Curriculum Committee	17	0	0	1/20/2017
Department of Computer Science	23	0	0	1/31/2017
LAS Curriculum Committee	7	0	0	2/17/2017
LAS Representative Assembly	18	0	0	2/28/2017
Faculty Senate Curriculum Committee	5	0	0	3/7/2017
Faculty Senate Academic Affairs Council	7	0	0	3/21/2017
Faculty Senate				

[FSCC – November 2013]