DATA SCIENCE

Data Science

Data Science is a rapidly growing academic discipline fueled by the proliferation of rich and complex data emerging from activities in science, industry, and governments. As a result, there is strong demand for data science professionals today in Iowa as well as across the nation and globe, and this market is expected to continue to grow in the next decade. The data science programs are intended for students who wish to study the data science discipline for its own sake as well as for students studying any discipline at Iowa State University with the goal of enabling them to work in data science. The courses in the data science program 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. The capstone will provide an opportunity for students to apply data science concepts to a domain problem while working in a multi-disciplinary team setting.

The Data Science major is intended for students with strong quantitative backgrounds and has the goal of educating students on the technical fundamentals of Data Science, with a focus on developing the knowledge and skills needed to transform data into insights. The major is an excellent opportunity for individuals who want to prepare themselves for the exciting Data Scientist positions that are in high demand today.

The minor in Data Science is intended for students studying any discipline at Iowa State and is designed to give students an in-depth understanding of data science as it is applied to a variety of domains.

The certificate in Data Science is intended for students studying any discipline at Iowa State and is designed to prepare them for future work with significant data science components. The capstone will provide an opportunity for students to apply data science concepts to a domain problem while working in a multi-disciplinary team setting.

Data Science Major

Purpose

This Bachelor's of Science degree program in Data Science is intended for students with strong quantitative backgrounds and has the goal of educating students on the technical fundamentals of data sciences, with a focus on developing the knowledge and skills needed to manage and analyze large-scale, heterogeneous data to address a wide range of problems.

Learning Outcomes

After successfully completing the program, students majoring in Data Science will demonstrate

1. an understanding of and an ability to apply the following data science concepts, tools and methods to data analysis pipelines:
   a. data acquisition
   b. data preprocessing
   c. exploratory data analysis
   d. inferential and predictive thinking, modeling and analysis
   e. computational thinking, data structures, and algorithms
2. an understanding of ethical, legal, societal, and economic concerns in application of data science concepts
3. an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders
4. an ability to function on multi-disciplinary teams using concepts and tools from data science

Requirements

The B.S. in Data Science consists of 120 total credit hours including: (1) 39 credits hours in the major core, three credits of which constitute a capstone course that is expected to provide experiential learning; (2) 9 credit hours in one of seven elective tracks to examine applications and theory of data sciences in a specific area; and (3) 23 credit hours of foundation courses. The capstone course will provide an opportunity for students to apply data science concepts to an application area while working in a multi-disciplinary team setting.

Data Science Major Requirements

Data Science Core Courses		39
DS 110	Orientation to Data Science	R
DS 201	Introduction to Data Science	3
DS 202	Data Acquisition and Exploratory Data Analysis	3
DS 303	Concepts and Applications of Machine Learning	3
DS 401	Data Science Capstone	3
COM S 228	Introduction to Data Structures	3
COM S 230 or CPR E 310	Discrete Computational Structures Theoretical Foundations of Computer Engineering	3
COM S 311	Introduction to the Design and Analysis of Algorithms	3
COM S 363	Introduction to Database Management Systems	3
CPR E 419	Software Tools for Large Scale Data Analysis	4
STAT 301	Intermediate Statistical Concepts and Methods	4
STAT 347	Probability and Statistical Theory for Data Science	4
STAT 457	Applied Categorical Data Analysis	3

At least 9 credits from any **ONE** of the following seven application emphasis areas:

Big Data		9-10
COM S 424	Introduction to High Performance Computing	3
COM S 426	Introduction to Parallel Algorithms and Programming	4
COM S 435	Algorithms for Large Data Sets: Theory and Practice	3
COM S 454	Distributed Systems	3
COM S 461	Principles and Internals of Database Systems	3
COM S 474	Introduction to Machine Learning	3
Engineering Applications		10
CPR E 388	Embedded Systems II: Mobile Platforms	4
CPR E 425	High Performance Computing for Scientific and Engineering Applications (cross-listed as COM S 425)	3
E E 425	Machine Learning: A Signal Perspective	3
Optimization		9
I E 312	Optimization	3
I E 483	Data Mining	3
I E 487	Big Data Analytics and Optimization	3
Security		9
COM S 421	Logic for Mathematics and Computer Science	3
COM S 453	Privacy Preserving Algorithms and Data Security	3
CPR E 431	Basics of Information System Security	3
Software Analytics		9
COM S 342	Principles of Programming Languages	3
COM S 413	Foundations and Applications of Program Analysis	3
COM S 440	Principles and Practice of Compiling	3
COM S 474	Introduction to Machine Learning	3
CPR E 416	Software Evolution and Maintenance	3
Statistics		9
STAT 402	Statistical Design and the Analysis of Experiments	3
STAT 407	Methods of Multivariate Analysis	3
STAT 421	Survey Sampling Techniques	3
COM S 474	Introduction to Machine Learning	3
Computational Biology		10
BCBIO 322	Introduction to Bioinformatics and Computational Biology	3
BCBIO 402	Fundamentals of Systems Biology and Network Science	3
BCBIO 444	Bioinformatic Analysis	4

Toward satisfying requirements of the College of Liberal Arts and Sciences, the following courses should be included:

COM S 227	Object-oriented Programming	4
MATH 165	Calculus I	4
MATH 166	Calculus II	4
MATH 265	Calculus III	4
MATH 207	Matrices and Linear Algebra	3
STAT 201	Introduction to Statistical Concepts and Methods	4
Foreign Language 3 years in high school or 1 year in college		0 - 8
Natural Science		8
Social Science		9
Arts and Humanities		12

The following courses meet the communication proficiency requirement:

LIB 160	Information Literacy	1
ENGL 150	Critical Thinking and Communication	3
ENGL 250	Written, Oral, Visual, and Electronic Composition	3
One of the following:		
ENGL 302	Business Communication	3
ENGL 314	Technical Communication	3
ENGL 332	Visual Communication of Quantitative Information (cross-listed as STAT 332)	3

According to the university-wide Communication Proficiency Grade Requirement, students must demonstrate their communication proficiency by earning a grade of C or better in ENGL 250. The Data Science program requires a C or higher in the upper-level ENGL course (302, 314, or 332).

All students must complete 3 credits of US Diversity and 3 credits of International Perspective courses.

To obtain a bachelor's degree from the College of Liberal Arts and Sciences, curriculum in liberal arts and sciences, a student must earn at least 45 credits at the 300 level or above taken at a four-year college. All such credits, including courses taken on a pass/not pass basis, may be used to meet this requirement.

B.S., Data Science

Freshman

Fall	Credits	Spring	Credits
DS 110		R COM S 228	3
ENGL 150		3 MATH 166	4
LIB 160		1 STAT 201	4
MATH 165		4 Arts and Humanities	3
COM S 227	4		

Social Science	3		
	15		14

Sophomore

Fall	**Credits**	**Spring**	**Credits**
ENGL 250	3	DS 202	3
DS 201	3	STAT 301	4
COM S 230 or CPR E 310	3	MATH 207	3
MATH 265	4	Social Science	3
Natural Science	4	Arts and Humanities	3
	17		16

Junior

Fall	**Credits**	**Spring**	**Credits**
ENGL 302 or 314	3	DS 303	3
STAT 347	4	STAT 457	3
COM S 311	3	CPR E 419	4
COM S 363	3	Arts and Humanities	3
Elective or Foreign Language	3-4	Elective or Foreign Language	3-4
	16-17		16-17

Senior

Fall	**Credits**	**Spring**	**Credits**
Major Elective	3	DS 401	3
Major Elective	3	Major Elective	3
Arts and Humanities	3	Social Science	3
Natural Science	4	Electives 300+	4-6
	13		13-15

The major elective courses will come from any one application emphasis area as outlined on the Undergraduate Major page. A student must take at least 9 credits from any single application emphasis area and may choose from: Big Data; Engineering Applications; Optimization; Security; Software Analytics; Statistics; and Computational Biology.

Data Science Minor

Purpose

The minor in data science is intended for students studying any discipline at Iowa State and is designed to give students an in-depth understanding of data science as it is applied to a variety of domains. 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.

Learning Outcomes

After completing the minor in data science, students will demonstrate:

- an ability to apply data science concepts, tools and technologies to data analysis pipelines,
- an understanding of ethical, legal, societal, and economic concerns in application of data science concepts,
- an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders, and
- an ability to function on multi-disciplinary teams using concepts and tools from data science.

Requirements

The minor in data science requires the completion of 15 credit hours, including 9 credits from the data science core and 6 credits from approved data science electives.

At least 6 credits must be taken in courses numbered at the 300-level or above.

At least 9 credits used for the minor cannot be used to meet any other department, college or university requirement for the baccalaureate degree except to satisfy the total credit requirement for graduation and to meet credit requirements in courses numbered 300 or above.

Courses for the minor cannot be taken on a pass/not-pass basis.

Course Requirements for Data Science Minor

Core Courses (9 credits)		
DS 201	Introduction to Data Science (Required)	3
DS 202	Data Acquisition and Exploratory Data Analysis (Required)	3
DS 301	Applied Data Modeling and Predictive Analysis (Required)	3
* DS 301 has a prerequisite of an introductory statistics course: STAT 101, STAT 104, STAT 105, STAT 201, STAT 226, STAT 231, STAT 305, STAT 322, or STAT 330.		
Electives (6 credits)		
A B E 316	Applied Numerical Methods for Agricultural and Biosystems Engineering	3
BCBIO 322	Introduction to Bioinformatics and Computational Biology	3
COM S 311	Introduction to the Design and Analysis of Algorithms	3
COM S 363	Introduction to Database Management Systems	3
COM S 424	Introduction to High Performance Computing	3
COM S 435	Algorithms for Large Data Sets: Theory and Practice	3
COM S 453X	Privacy Preserving Algorithms and Data Security	3
COM S 474	Introduction to Machine Learning	3

C R P 251X	Introduction to Geographic Information Systems	3
C R P 351X	Intermediate Geographic Information Systems	3
C R P 452	Geographic Data Management and Planning Analysis	3
C R P 456	GIS Programming and Automation	3
CPR E 419	Software Tools for Large Scale Data Analysis	4
CPR E 426	Introduction to Parallel Algorithms and Programming	4
ECON 371	Introductory Econometrics	4
ENGL 332	Visual Communication of Quantitative Information	3
FIN 450X	Analytical Finance	3
I E 312	Optimization	3
I E 483	Data Mining	3
LING 410X	Language as Data	3
MIS 436	Introduction to Business Analytics	3
MIS 446	Advanced Business Analytics	3
MKT 368	Marketing Analytics	3
STAT 301	Intermediate Statistical Concepts and Methods	4
STAT 330	Probability and Statistics for Computer Science	3
STAT 407	Methods of Multivariate Analysis	3
STAT 430	Empirical Methods for the Computational Sciences	3
STAT 457	Applied Categorical Data Analysis	3
STAT 480	Statistical Computing Applications	3

Data Science Certificate

Purpose

The certificate in data science is intended for students studying any discipline at Iowa State and is designed to prepare them for future work with significant data science components. The data science certificate is also available to students who have already earned a Baccalaureate degree from Iowa State or elsewhere. The capstone will provide an opportunity for students to apply data science concepts to a domain problem while working in a multi-disciplinary team setting. The certificate in data science will prepare students with the technical and communication skills to enter the workforce as domain experts with data science skills.

Learning Outcomes

After completing the certificate in data science, students will demonstrate:

- an ability to apply data science concepts, tools and technologies to data analysis pipelines,
- an understanding of ethical, legal, societal, and economic concerns in application of data science concepts,
- an ability to visualize, interpret and communicate the output of data analysis pipelines to stakeholders, and
- an ability to function on multi-disciplinary teams using concepts and tools from data science.

Requirements

The certificate in data science requires the completion of 21 credit hours, including 9 credits from the data science core, 9 credits from approved data science electives, and a three-credit data science capstone experience.

At least 9 credits must be taken in courses numbered at the 300-level or above.

At least 9 credits used for the certificate cannot be used to meet any other department, college or university requirement for the baccalaureate degree except to satisfy the total credit requirement for graduation and to meet credit requirements in courses numbered 300 or above.

Courses for the certificate cannot be taken on a pass/not-pass basis.

Course Requirements for Data Science Certificate

Core Courses (9 credits)		
DS 201	Introduction to Data Science (Required)	3
DS 202	Data Acquisition and Exploratory Data Analysis (Required)	3
DS 301	Applied Data Modeling and Predictive Analysis (Required)	3
* DS 301 has a prerequisite of an introductory statistics course: STAT 101, STAT 104, STAT 105, STAT 201, STAT 226, STAT 231, STAT 305, STAT 322, or STAT 330.		
Electives (9 credits)		
A B E 316	Applied Numerical Methods for Agricultural and Biosystems Engineering	3
BCBIO 322	Introduction to Bioinformatics and Computational Biology	3
COM S 311	Introduction to the Design and Analysis of Algorithms	3
COM S 363	Introduction to Database Management Systems	3
COM S 424	Introduction to High Performance Computing	3
COM S 435	Algorithms for Large Data Sets: Theory and Practice	3
COM S 453X	Privacy Preserving Algorithms and Data Security	3
COM S 474	Introduction to Machine Learning	3
C R P 251X	Introduction to Geographic Information Systems	3
C R P 351X	Intermediate Geographic Information Systems	3

C R P 452	Geographic Data Management and Planning Analysis	3
C R P 456	GIS Programming and Automation	3
CPR E 419	Software Tools for Large Scale Data Analysis	4
CPR E 426	Introduction to Parallel Algorithms and Programming	4
ECON 371	Introductory Econometrics	4
ENGL 332	Visual Communication of Quantitative Information	3
FIN 450X	Analytical Finance	3
I E 312	Optimization	3
I E 483	Data Mining	3
LING 410X	Language as Data	3
MIS 436	Introduction to Business Analytics	3
MIS 446	Advanced Business Analytics	3
MKT 368	Marketing Analytics	3
STAT 301	Intermediate Statistical Concepts and Methods	4
STAT 330	Probability and Statistics for Computer Science	3
STAT 407	Methods of Multivariate Analysis	3
STAT 430	Empirical Methods for the Computational Sciences	3
STAT 457	Applied Categorical Data Analysis	3
STAT 480	Statistical Computing Applications	3
Data Science capstone experience (3 credits)		
DS 401	Data Science Capstone	3

Courses primarily for undergraduates:

DS 110: Orientation to Data Science

Cr. R.

Introduction to the procedures and policies of Iowa State University and the Data Science program, test-outs, honorary societies, etc. Issues relevant to student adjustment to college life will also be discussed. Offered on a satisfactory-fail basis only.

DS 201: Introduction to Data Science

Cr. 3. Alt. F., offered irregularly.Alt. S., offered irregularly.

Prereq: 1-1/2 Years of High School Algebra

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.

DS 202: Data Acquisition and Exploratory Data Analysis

Cr. 3. Alt. F., offered irregularly.Alt. S., offered irregularly.

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. Programming exercises.

DS 301: Applied Data Modeling and Predictive Analysis

Cr. 3. Alt. F., offered irregularly.Alt. S., offered irregularly.

Prereq: DS 201, 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. Participation in a multi-disciplinary team project.

DS 303: Concepts and Applications of Machine Learning

Cr. 3.

Prereq: DS 201

Machine learning concepts such as training and test sets; feature extraction; principles of machine learning techniques; regression; pattern recognition methods; unsupervised learning techniques; assessment and diagnostics: overfitting, error rates, residual analysis, model assumptions checking, feature selection; ethical issues in data science; communicating findings to stakeholders in written, oral, visual and electronic form.

DS 401: Data Science Capstone

Cr. 3. Alt. F., offered irregularly.Alt. S., offered irregularly.

Prereq: DS 202X; DS 301X

Students work as individuals and teams to complete the planning, design, and implementation of a significant multi-disciplinary project in data science. Oral and written reports.