Data Science and Analytics

Data science is a rapidly blossoming field of study and career with a highly multidisciplinary characteristic. The confluence of big data, massively powerful cloud computing platforms, and need of businesses from all sectors to leverage their data repositories has created a high-growth environment and demand for data scientists. Data scientists routinely leverage tools and techniques from computer science, information systems, advanced statistics, and machine learning. To satisfy the growing need for data scientist who can transform large collections of data into actionable decision making products for their employers, we are proposing the Master of Science in Data Science and Analytics.

This multidisciplinary Data Science and Analytics (DSA) degree program will consist of 34-credit hours of learning in the online and mixed mode format in which students will visit campus one time each academic year for an intensive on site learning experience.

The academic program will consist of 19 credit hours of core, fundamental data science courses; followed by 9 credits of emphasis area specific courses and 6 credits of industry relevant case studies and capstone project courses.

Data Science is an emerging discipline that, by its nature, integrates traditional disciplines. The proposed degree program will leverage prior investments in the computing disciplines across campuses and colleges within each campus. The MU Informatics Institute will coordinate this collaborative degree program by leveraging existing courses from Computer Science, Journalism, and Information Science & Learning Technologies Departments to deliver the various core and emphasis area course. Existing courses will be adapted to the online format, and new courses that are properly focused and structured for the DSA program will be developed.

Faculty

**Associate Professor** S. Goggins

**Assistant Research Professor** G. Scott

* Graduate Faculty Member - membership is required to teach graduate-level courses, chair master's thesis committees, and serve on doctoral examination and dissertation committees.

** Doctoral Faculty Member - membership is required to chair doctoral examination or dissertation committees. Graduate faculty membership is a prerequisite for Doctoral faculty membership.

Undergraduate

While MU does not offer undergraduate degrees specifically in Data Science and Analytics, the University does offer baccalaureate opportunities in a number of related areas.

A listing of current degree programs can be found here (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics)

Graduate

* MS in Data Science and Analytics (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/ms-data-science-analytics)

* with emphasis in Biotechnology (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/ms-data-science-analytics-emphasis-biotech)

* with emphasis in High Performance Computing (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/ms-data-science-analytics-emphasis-high-perform-comp)

* with emphasis in Strategic Communications and Data Journalism (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/ms-data-science-analytics-emphasis-strat-com-data-journ)

* with emphasis in Human Centered Data (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/ms-data-science-analytics-emphasis-human-centered-data)

* Graduate Certificate in Data Science and Analytics (http://catalog.missouri.edu/undergraduategraduate/interdisciplinaryacademicprograms/data-science-analytics/grad-cert-data-sci-analytics)

DATA_SCI 7600: Introduction to Data Science and Analytics

An introductory course in data science and analytics. The objective of the course is to give students a broad overview of the various aspects of data analytics such as accessing, cleansing, modeling, visualizing, and interpreting data. Students will perform hands-on learning of data analytic topics, using technologies such as Python, R, and open source analytic tools. Two Big Data cyberinfrastructure platforms will be introduced through case studies, allowing students to perform data analytical learning modules on modern cloud infrastructure and other relevant technologies. Graded on A-F basis only.

**Credit Hours:** 3

**Recommended:** Basic programming experience and Basic database experience

DATA_SCI 7601: Introduction to Data Science

This course is an introduction to the NGA Program of Study in Data Science (PSDS), the concentration areas, and the role of each concentration area in data science. Participants will learn how to receive, if desired, an accredited Graduate Certificate and/or a Master of Science degree in Data Science and Analytics from the University of Missouri. Participants will receive an introduction to software, tools, and resources to be utilized throughout the program. Participants will learn of systematic methodologies for data science projects and the data science pipeline through review of case studies. Graded on S/U basis only.

**Credit Hours:** 2

**Recommended:** Enrollment in NGA Training Program or instructor consent

DATA_SCI 7610: Python Programming Boot Camp

This course teaches students how to program in Python, including use of auxiliary libraries various Python ecosystems. Students are introduced to the iPython notebooks from the SciPy ecosystem, as well Python’s use across the spectrum of Data Science courses and topics. Many activities are focus on data ingestion, cleaning, manipulation, and restructuring (e.g., ETL). Graded on A-F basis only.

**Credit Hour:** 1

**Recommended:** Enrollment in NGA Training Program or instructor consent
DATA_SCI 7620: Database Basics and SQL Boot Camp
This course covers a core concepts to heterogeneous data management, including relational databases, NoSQL databases, and other data storage systems. The focus is on making students quickly productive in the use of multiple types of database management systems available on the market for data science work. This includes traditional relational databases, NOSQL databases and graph databases. This course is a 1 credit hour / 5 day course. This course is delivered in an asynchronous online mode. The instructor virtually kicks off the course on day-one, then four additional days over a two week period are used for self-paced, online activities using the JupyterHub learning environment. Graded on A-F basis only.

Credit Hours: 1
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 7630: Introductory Probability and Statistics for Data Analytics
This course explores the use of inferential and predictive statistics for data modeling and analytics. Single-variate and multivariate statistical concepts are discussed, as well as intermediate exposure to statistical modeling. Students learn to evaluate model effectiveness and conduct results driven model selection. Statistical and modeling techniques focus on high dimensional data analytics. Topics related to dimensionality reduction are also covered, such as principal component analysis and factor analysis. Graded on A-F basis only.

Credit Hours: 2
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 7640: R Statistical Programming Boot Camp
This course teaches students how to program in R, including use of auxiliary libraries in R focused on various statistical and visualization oriented techniques. Students are introduced to R's use across the spectrum of Data Science courses and topics. Many activities focus on the development of statistical tests, and the use of R for statistical exploration. This course teaches students how to program in R, including use of auxiliary libraries in R focused on various statistical and visualization oriented techniques. Students are introduced to R's use across the spectrum of Data Science courses and topics. Many activities focus on the development of statistical tests, and the use of R for statistical exploration. Graded on A-F basis only.

Credit Hours: 1
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 8610: Statistical and Mathematical Foundations for Data Analytics
An intermediate statistics class designed to build the mathematical foundation for students dealing with Big Data phenomena. Topics include discussions of probability, data sampling, data summarization, sampling distributions, statistical inference, statistical pattern analysis, hypothesis testing, regression, and nonparametric inference over multidimensional data collections. Students will engage in Big Data projects using various publicly available data sets and leveraging modern Data Science tools, techniques, and cyberinfrastructure. Graded on A-F basis only.

Credit Hours: 3
Recommended: Basic understanding of mathematical principles of vectors and matrices, and Basic course in probability and statistics

DATA_SCI 8612: Spatial and Geostatistical Analysis
This course will provide a practical overview of key issues encountered when working with and analyzing spatial data as well as an overview of major spatial analysis approaches. Discussions and laboratory work will focus on implementation, analysis, and interpretive issues given constraining factors that commonly arise in practice. Graded on A-F basis only.

Credit Hours: 3
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 8614: Data Analytics from Applied Machine Learning
This course leverages the foundations in statistics and modeling to teach applied concepts in machine learning. Participants will learn various classes of machine learning and modeling techniques, and gain an in-depth understanding how to select appropriate techniques for various data science tasks. Topics cover a spectrum from simple Bayesian modeling to more advanced algorithms such as support vector machines, decision trees/forests, and neural networks. Students learn to incorporate machine learning workflows into data-intensive analytical processes. Graded on A-F basis only.

Credit Hours: 3
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 8620: Database and Analytics
Covers the Fundamental concepts of current database systems and query methods with emphasis on relational model and non-relational techniques in Big Data environments. Topics include entity-relationship model, relational algebra, indexing, query optimization, normal forms, tuning, security, NoSQL, and data analytics skills in both relational and non-relational environments. Project work involves modern relational DBMS systems and NoSQL environments. Graded on A-F basis only.

Credit Hours: 3
Recommended: Basic understanding of mathematical principles of vectors and matrices, and Basic course in probability and statistics

DATA_SCI 8630: Data Mining and Information Retrieval
The course introduces the main concepts and techniques of data mining and information retrieval. It covers a variety of data mining topics and methods to extract hidden and predictive patterns from large data collections. Furthermore, theory and techniques for the modeling, indexing, and retrieval of relational, nonrelational, text-based and multimedia databases is covered. Topics include introduction to data mining process, mining frequent patterns, and pattern analysis, as well as different information retrieval models and evaluation, query languages and operations, and indexing/searching methods. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7600 and DATA_SCI 8620
Recommended: Basic understanding of mathematical principles of vectors and matrices; Basic course in probability and statistics; Basic course in databases and data analytics
DATA_SCI 8635: Cloud Computing for Data Analytics
This course introduces students to cluster and cloud computing big data ecosystems. Topics include a survey of cloud computing platforms, architectures, and use-cases. Students will examine scaling data science techniques and algorithms using a variety of cluster and cloud paradigms, such as those built atop Hadoop (Map-Reduce) concepts, and others. Graded on A-F basis only.
Credit Hours: 3
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 8640: Big Data Security
This course provides an overview of state-of-the-art topics in Big Data Security, looking at data collection (smartphones, sensors, the Web), data storage and processing (scalable relational databases, Hadoop, Spark, etc.), extracting structured data from unstructured data, systems issues (exploiting multicore, security), Securing sensitive data, personal data and behavioral data while ensuring a respect for privacy will be a focus point in the course Graded on A-F only.
Credit Hours: 3
Prerequisites: DATA_SCI 7600 and DATA_SCI 8620

DATA_SCI 8650: Big Data Visualization
Covers the Fundamental concepts of current visualization concepts and technologies. Unlike many data visualization courses, this one focuses on principles of visualization design and the grammar of graphics. These principles are then implemented in popular contemporary visualization technologies. Students will develop an advanced knowledge of the appropriate selection, modeling, and evaluation of data visualizations. Graded on A-F basis only.
Credit Hours: 3
Prerequisites: DATA_SCI 7600 and DATA_SCI 8620
Recommended: Basic understanding of mathematical principles of vectors and matrices; Basic course in probability and statistics; Basic course in databases and data analytics

DATA_SCI 8654: Advanced Visualization and Communication I
Covers the Fundamental concepts of current visualization concepts and technologies, adding in Infographic and Interactive Visualization Design. Unlike many data visualization courses, this one focuses on principles of visualization design and the grammar of graphics as they can be applied to combining art & technology to tell data stories. These principles are then implemented in popular contemporary visualization technologies. Students will develop an advanced knowledge of the appropriate selection, modeling, and evaluation of data visualizations. Graded on A-F basis only.
Credit Hours: 3
Recommended: Enrollment in NGA Training Program or instructor consent

DATA_SCI 8656: Advanced Visualization and Communication II
Covers the Fundamental concepts of current visualization concepts and technologies, adding in Infographic and Interactive Visualization Design. Unlike many data visualization courses, this one focuses on principles of visualization design and the grammar of graphics as they can be applied to combining art & technology to tell data stories. These principles are then implemented in popular contemporary visualization technologies. Students will develop an advanced knowledge of the appropriate selection, modeling, and evaluation of data visualizations. Graded on A-F basis only.
Credit Hours: 3
Prerequisites: DATA_SCI 8610, DATA_SCI 8620

DATA_SCI 8660: Data and Information Ethics
Introduces the ethics related to Big Data in industry, business, academia, and research settings. Students will learn the social, ethical, legal and policy issues that underpin the big data phenomenon. Discussions and case studies will help guard against the repetition of known mistakes and inadequate preparation. The course content will follow the guidelines to be developed by the Council for Big Data, Ethics, and Society. Graded on A-F basis only.
Credit Hour: 1
Prerequisites: DATA_SCI 7600 and DATA_SCI 8650

DATA_SCI 8680: Big Data Analysis Case Study
Using a case-study approach, students will engage in discussions on a variety of big data topics relevant to their emphasis area and the realm of Big Data. This course will help students generate ideas and prepare them for the Big Data Capstone. Course work will be performed in small teams, mentored by faculty and/or industry advisors. Teams will research, cultivate, curate, and leverage large data sets. Students will gain hands-on experience applying relevant data science and analytical technology and techniques to gain insight and information from these real-world data sets. Graded on A-F basis only.
Credit Hours: 3
Prerequisites: DATA_SCI 8630, DATA_SCI 8640, DATA_SCI 8650

DATA_SCI 8750: Parallel Computing for Data Science
This course will provide in-depth treatment of the evolution of high performance, parallel computing architectures and how these architectures and computational ecosystems support data science. We will cover topics such as: parallel algorithms for numerical processing, parallel data search, and other parallel computing algorithms which facilitate advanced analytics. To reinforce lecture topics, learning activities will be completed using parallel computing techniques for modern multicore and multi-node systems. Parallel algorithms will be investigated, selected, and then developed for various scientific data analytics problems. Programming projects will be completed using Python and R, leveraging various parallel and distributed computing infrastructure such as AWS Elastic Map Reduce and Google Big Query. Students will research emerging parallel and scalable architectures for data analytics. Graded on A-F basis only.
Credit Hours: 3
Prerequisites: DATA_SCI 8610, DATA_SCI 8620