

# DATA SCIENCE, MS

## Program Description

A&M-Corpus Christi's Data Science program will prepare students to meet the growing state, national, and international needs for highly qualified personnel in the fields of data science. The program objectives underline the interdisciplinarity of data science and the importance of building a strong foundation of data science for our students.

- Provide strong core training so that graduates can adapt easily to changes and new demands from society and industry.
- Develop an in-depth understanding of the theory and methods in data science and develop students' skills for problem analysis and decision-making.
- Integrate fields within computer science, optimization, engineering, and statistics to create adept and well-rounded data scientists.
- To teach students how to create new methodologies and application tools to solve interdisciplinary problems with Big Data.
- Enable students to communicate effectively how to resolve problems from big, complex, and unstructured data.
- Provide students with insights into data science practice in interdisciplinary fields.

## Admission Requirements

Persons seeking admission to the M.S. in Data Science should first contact the program faculty and identify a faculty member willing to serve as the graduate adviser. Applicants will not be admitted to the program without a graduate adviser.

Applicants for the M.S. in Data Science should have the equivalent of a bachelor's in data science or other areas of science, with the equivalent of at least a minor in Mathematics or Statistics. Specific leveling course work is MATH 1442 Statistics for Life, MATH 3311 Linear Algebra, MATH 3315 Differential Equations, MATH 3342 Applied Probability and Statistics, MATH 2415 Calculus III, and MATH 4301 Introduction to Analysis. Students with no computer programming experience may find themselves at a disadvantage in certain courses without an introductory programming course.

## Program Requirements

- Students may enter with a BS or an MS.
- Adequate preparation with coursework in mathematics, statistics, and computer science is required (or leveling courses).
- Mathematics: discrete mathematics, calculus sequence, linear algebra, numerical methods, differential equations
- Statistics: probability theory, advanced statistics beyond introductory courses, mathematical statistics
- Computer science: high-level programming language (python, SQL, C++ or equivalent), MATLAB, R-statistical package, Data Structures
- All students take fundamental core courses on Data Science.
- Depending on individual interest and preferences, students select their own emphasis/track for future studies, research, and dissertation.
- Diverse program faculty could offer tracks with accents on Environmental and Marine Sciences, Biology, Engineering, Business, Health Care, etc.

- To successfully graduate with MS degree, students must complete course work and defend a thesis: total 30 SCH.
- Core courses are a total of 12 SCH.
- Prescribed elective courses are 12 SCH.
- Thesis research and thesis are 6 SCH.

Code	Title	Hours
<b>Core Courses</b>		
DASC 5301	Principles of Data Science	3
DASC 5302	Data Science and Predictive Analytics	3
DASC 5307	Machine Learning in Data Science	3
or COSC 6338	Machine Learning	
DASC 5323	Natural System Analysis and Multivariate Statistics	3
or CMSS 6303	Natural Systems Analysis	
<b>Electives</b>		
Select 12 hours from the following:		12
DASC 5303	Data Science Computing	
DASC 5304	Bayesian Interference in Data Science	
or MATH 6318	An Introduction to Bayesian Statistics	
DASC 5305	Applied Differential Equations in Data Science	
DASC 5308	Numerical Methods for Data Science	
DASC 5306	Dynamical System Analysis for Data Science	
DASC 5321	Geospatial Data Structure	
or GSCS 6321	Geospatial Data Structures	
DASC 5324	Digital Image Processing	
or COSC 6321	Digital Image Processing	
DASC 5325	Natural Systems Modeling	
or CMSS 6305	Natural Systems Modeling	
DASC 5327	Introduction to Computer Graphic/COSC 6327	
DASC 5329	Scientific Visualization/GSCS 6329/GSEN 6329	
DASC 5331	Advanced Geospatial Computing	
or GSCS 6331	Advanced Geospatial Computing	
DASC 5336	Database Management Systems	
or COSC 6336	Database Management Systems	
DASC 5337	Data Mining/COSC 6337	
DASC 5340	Genomics, Proteomics and Bioinformatics	
or BIOL 5340	Genomics, Proteomics and Bioinformatics	
DASC 5341	Statistical Methods and Data Analysis	
or MATH 5341	Statistical Methods and Data Analysis	
DASC 5342	Linear Statistical Models	
or MATH 5342	Linear Statistical Models	
DASC 5345	Computational Methods for Statistics	
or MATH 5345	Computational Methods for Statistics	
DASC 5348	Optimization	
or MATH 5348	Optimization	
DASC 5350	Advanced Topics in DBMS	
or COSC 6350	Advanced Topics in DBMS	
DASC 5352	Environmental Forecasting	
or CMSS 6352	Environmental Forecasting	
DASC 5354	Artificial Intelligence	
or COSC 6354	Artificial Intelligence	
DASC 5355	Data Communication and Networking	

	or COSC 6355 Data Communications and Networking	
DASC 5356	Computational Biology	
DASC 5365	Spatial Database Design	
	or COSC 6365 Current Trends in Programming	
DASC 5380	Data Analytics	
	or COSC 6380 Data Analytics	
DASC 5386	Remote Sensing and Image Analysis	
	or GSEN 6386 Remote Sensing and Image Analysis	
<b>Thesis Option</b>		
DASC 5994	Proposal Research	3
DASC 5995	Thesis	3
<b>Total Hours</b>		<b>30</b>

## Courses

### DASC 5301 Principles of Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

This course integrates skill set spanning mathematics, statistics, machine learning, databases, and computer science along with a good understanding of the craft of problem formulations in STEM fields to find effective solutions. This course will introduce basic principles and tools in data science and will expose students to concepts and techniques to deal with various facets of data science practice, including data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, and evaluation. R and other statistical software will be used to make the learning contextual.

### DASC 5302 Data Science and Predictive Analytics

#### 3 Semester Credit Hours (3 Lecture Hours)

This course aims to build computational abilities, inferential thinking, and practical skills to solve complex problems in data science and make predictive models. It uses the concepts in data management, statistical modeling, statistical computing, and visualization, and integrates the use of programming in R or similar languages to analyze, model, analyze, and interpret large, multi-source heterogeneous data.

### DASC 5303 Data Science Computing

#### 3 Semester Credit Hours (3 Lecture Hours)

Programming and computing techniques for the requirements of data science: acquisition and organization of data; visualization, modeling, and inference for scientific applications; presentation and interactive communication of results. Emphasis on computing for substantial projects. Software development with an emphasis on R, plus other key software tools.

### DASC 5304 Bayesian Inference in Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

This course introduces the Bayesian approach. It involves the concept of probability and the analysis of data which focuses on the principles of data analysis and computer-intensive, modern statistical modeling. Topics include Bayesian inference, prior and posterior distributions, regression modeling, hierarchical models, model checking and selection, missing data, and stochastic simulation by Markov Chain Monte Carlo including Gibbs sampling and Metropolis algorithms. The course will apply Bayesian methods to practical problems, by building models from the prior probabilities to the posterior distribution with statistical packages.

### DASC 5305 Applied Differential Equations in Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

The purpose of this course is to study real-world applications of differential equations to problems in Data Science. The course will cover how differential equations can be used in the development, solution, and analysis of mathematical models based on data. The course will include learning techniques for using data to estimate the parameters and structure of a model and learning about evaluation techniques to determine whether a particular model is a good one.

### DASC 5306 Dynamical System Analysis for Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

The purpose of this course is to study the modern perspective on Data-driven Dynamical Systems. Specifically, we will focus on the key challenges of discovering dynamics from data and finding data-driven representations that make nonlinear systems amenable to linear analysis. Dynamic mode decomposition, Koopman operators, diffusion maps, equations free modeling, Lagrangian coherent systems, finite-time Lyapunov exponents — are some of the new methods that have been introduced in recent decades to analyze dynamical systems. The lectures will survey these methods along with earlier ones of “nonlinear time series analysis.” The goal will be to describe theoretical principles and algorithmic approaches suitable for working with empirical data and computer defined systems.

### DASC 5307 Machine Learning in Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

Machine learning is a highly interdisciplinary subject that encompasses the techniques from statistics, probability, linear algebra, optimization, and computer science. Machine learning techniques are being used in several areas such as face recognition, self-driving cars, cybersecurity, and also in the areas where decisions are very important without human intervention. This course covers both theory and practical algorithms for machine learning for a variety of applications. We cover topics such as supervised learning (generative/discriminative learning, parametric/nonparametric learning, neural networks, and support vector machines), unsupervised learning (clustering, dimension reduction, kernel methods), learning theory, reinforcement learning, and adaptive control. This course will also discuss a variety of recent applications of machine learning, such as data mining, autonomous navigation, and web data processing (maybe from Facebook or Twitter).

### DASC 5308 Numerical Methods for Data Science

#### 3 Semester Credit Hours (3 Lecture Hours)

This course introduces a broad range of numerical techniques that are widely used in mathematics, science, and engineering. The topics covered in this course include basic direct and iterative methods for linear systems; classical root-finding methods; Newton's method and related methods for nonlinear systems; fixed-point iteration; polynomial, piecewise polynomial, and spline interpolation methods; least-squares approximation; orthogonal functions and approximation; basic techniques for numerical differentiation; numerical integration, including adaptive quadrature; and methods for initial-value problems for ordinary differential equations. Python, or similar software, will be used to implement the methods covered in the class.

**DASC 5311 Statistical Learning****3 Semester Credit Hours (3 Lecture Hours)**

This course will introduce students the problem of supervised (classification and regression) and unsupervised learning (dimension reduction and clustering) from the perspective of statistical learning. It aims to go far beyond the classical statistical methods. Students will learn a collection of flexible tools and techniques for using data to construct prediction algorithms and perform data analysis. Topics will include splines & generalized additive models, model selection & regularization methods (ridge and lasso), tree-based methods, random forests & boosting as well as classical linear approaches such as Logistic Regression, Linear Discriminant Analysis, K-Means, Clustering and Nearest Neighbors. Programming in R will be used to provide hands-on training and examples.

**DASC 5321 Geospatial Data Structure****3 Semester Credit Hours (3 Lecture Hours)**

The representation of spatial data is an important issue in diverse areas including computer graphics, geographic information systems (GIS), robotics, and many others. Choosing an appropriate representation is a key to facilitate operations such as spatial search. This course will focus on representation of point data and object data, which are the important types of spatial data. Various fundamental data structures on spatial data, such as quadtrees, kd-trees, grid structures, kd-trees, and R-trees will be explored. The use of these structures to address some important problems will also be covered.

**DASC 5323 Natural System Analysis and Multivariate Statistics****3 Semester Credit Hours (3 Lecture Hours)**

Statistical analysis for data collected in several variables. Topics include sampling from multivariate normal distribution, multivariate analysis of variance, discriminant analysis, principle components, and factor analysis.

**DASC 5324 Digital Image Processing****3 Semester Credit Hours (3 Lecture Hours)**

This course introduces concepts and techniques for image processing. The objective of this course is to introduce the fundamental techniques and algorithms used for processing and extracting useful information from digital images. The students will learn how to apply the image processing methods to solve real-world problems.

**DASC 5325 Natural Systems Modeling****3 Semester Credit Hours (3 Lecture Hours)**

Modeling and analysis of deterministic and stochastic dynamical systems, including investigation of model behavior and stability. Theory will be applied to research natural environmental and biological systems such as multi-species systems, epidemic models, carbon circulation in the biosphere, Nutrients- Phytoplankton-Zooplankton models, etc.

**DASC 5328 Data and Information Visualization****3 Semester Credit Hours (3 Lecture Hours)**

This course focuses on building creative and technical skills to transform data into visual reports for the purpose of engendering a shared understanding. Students will learn to use software to ingest, organize, and visualize data, with an emphasis on applying design principles to produce clear, elegant graphs and dashboards that capture the essence of an insight, message, or recommendation distilled from the data.

**DASC 5331 Advanced Geospatial Computing****3 Semester Credit Hours (3 Lecture Hours)**

Seminar in reading and critical evaluation of academic literature in the field of and fields relating to geospatial computing. Student will design, implement, and evaluate an advanced, contemporary geospatial computing technology to solve a geospatial problem.

**DASC 5336 Database Management Systems****3 Semester Credit Hours (3 Lecture Hours)**

A study of contemporary database management concepts. Performance (indexing, query optimization, update optimization), concurrency, security and recovery issues are discussed. Also includes the study of front-end environments that access the database.

**DASC 5340 Genomics, Proteomics and Bioinformatics****3 Semester Credit Hours (3 Lecture Hours)**

Integrative biological study using genome-wide approaches and bioinformatics. The “-omics” technologies (Genomics, Proteomics, Metabolomics, etc.) will be reviewed. Applications to understanding biological function in various biological disciplines will be emphasized.

**DASC 5341 Statistical Methods and Data Analysis****3 Semester Credit Hours (3 Lecture Hours)**

Introduction to the basic concepts of probability, common distributions, statistical methods, data analysis and a wide variety of statistical inference techniques. Demonstrations of the interplay between probability models and statistical inference. Data sets will be analyzed using the R software package.

**DASC 5342 Linear Statistical Models****3 Semester Credit Hours (3 Lecture Hours)**

Review of basic concepts in probability theory. Principles of estimation and model building. Linear models, especially ANOVA and regression. Non-parametric alternatives.

**DASC 5345 Computational Methods for Statistics****3 Semester Credit Hours (3 Lecture Hours)**

An introduction to computing tools needed by the modern statistician. Topics include floating- point numbers, reformatting large datasets, important statistical algorithms, and parallel processing.

**DASC 5348 Optimization****3 Semester Credit Hours (3 Lecture Hours)**

Unconstrained optimization, necessary and sufficient conditions for solutions, basic algorithms. Constrained optimization, KKT conditions, linear programming, convex programming, algorithms.

**DASC 5350 Advanced Topics in DBMS****3 Semester Credit Hours (3 Lecture Hours)**

The study of emerging database technologies. Topics are chosen from data warehousing, distributed databases, spatial databases and web-based applications.

**DASC 5352 Environmental Forecasting****3 Semester Credit Hours (3 Lecture Hours)**

Statistical techniques (classic and Bayesian) and new artificial intelligence-based techniques, such as neural networks, for the analysis of environmental systems with large datasets.

**DASC 5354 Artificial Intelligence****3 Semester Credit Hours (3 Lecture Hours)**

Fundamental concepts and techniques for the design of computer-based, intelligent systems. Topics include: a brief history, methods for knowledge representation, heuristic search techniques, programming in LISP or Prolog.

**DASC 5355 Data Communication and Networking****3 Semester Credit Hours (3 Lecture Hours)**

Areas studied include principles of computer-based communication systems, analysis and design of computer networks, and distributed data processing.

**DASC 5356 Computational Biology****3 Semester Credit Hours (3 Lecture Hours)**

Introduces the powerful open-source computing tools that are used in biological research for the creation, organization, manipulation, processing, analysis, and archiving of "big data." This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research projects. The primary topics covered are data formats and repositories, command line Linux computing and scripting, regular expressions, super-computing, computer programming with PYTHON and R, data visualization with R, version control and dissemination of scripts and programs with GIT, and typesetting with markdown languages.

**DASC 5365 Spatial Database Design****3 Semester Credit Hours (3 Lecture Hours)**

This course will focus on spatial database principles and the practical skills of design, implementation, and use of spatial databases. This course will first cover fundamentals of relational database design, and then focus on design and management of spatial databases utilizing geodatabase models. In addition, case studies of geodatabase design models in several applications will also be covered. This course is intended for students who want to design, create, maintain and manipulate data from a geospatial database.

**DASC 5380 Data Analytics****3 Semester Credit Hours (3 Lecture Hours)**

This course will introduce state-of-the-art techniques to process and analyze different types of data, generate insights and knowledge from data, and make data-based decisions and predictions. Real-world examples will be used to familiarize students with the theory and applications. Main topics include data preprocessing, probability theory, tests of hypothesis, and various data analysis techniques (e.g., clustering, classification, prediction/forecasting, etc.) for different types of data including static, time-series, spatial, and spatiotemporal.

**DASC 5383 Advanced Geospatial Analytics****3 Semester Credit Hours (3 Lecture Hours)**

This course will focus on the theory, techniques, and applications of advanced geospatial analytics. Topics covered include spatial point patterns, network analysis, area objects and spatial autocorrelation, and spatial interpolation. New approaches to geospatial analytics will also be covered. This course emphasizes the methods and the applied side of geospatial analytics that can be useful in students' own theses or projects for their current or potential employers.

**DASC 5386 Remote Sensing and Image Analysis****3 Semester Credit Hours (3 Lecture Hours)**

Addresses the interpretation, processing and analysis techniques of remotely sensed data acquired by orbital and sub-orbital platforms. Physical principles and imaging mechanisms, remote sensing systems, data characteristics, image processing, and information extraction methods will be covered. Topics include passive optical imaging with multispectral, hyperspectral, and thermal sensing; active imaging with radar sensing; image corrections and rectification; spatial/frequency transforms and image filtering; image classification and feature extraction; and image processing with machine learning techniques. Applications in the course will be focused on geomatics and monitoring of natural and built environments.

**DASC 5390 Special Topics****3 Semester Credit Hours (3 Lecture Hours)**

An advanced study of a Data Science topic. May be repeated with full credit in another area of data science, statistics, or mathematics. Topics vary by semester and offering.

**DASC 5994 Proposal Research****3 Semester Credit Hours**

This course develops an ability to independently investigate a technical topic of interest, and the skills necessary to successfully communicate on that topic. The student learns how to find, organize, assimilate, and report on technical information derived from published sources. Specific areas of study include literature searches, technical word processing, technical writing style, and oral presentation techniques. A final paper and a formal presentation are submitted in lieu of a final exam in the final semester.

**DASC 5995 Thesis****3 Semester Credit Hours**

Students work with an advisor to complete and present their proposed thesis. Students may register for 3 to 9 semester hours per semester. Only 3 hours total will count toward the MS degree in data science.

**DASC 5997 Capstone Project****3-9 Semester Credit Hours**

This course develops an ability to independently investigate a technical topic of interest, and the skills necessary to successfully communicate on that topic. The student will work on the real world situation with industries, healthcare providers, environmental agencies and other entities in the need to work on big data and/or statistical/mathematical modeling. A final paper and a formal presentation are submitted in lieu of a final exam in the final semester before graduation.