

# MASTER OF SCIENCE IN DATA SCIENCE

<https://engineering.pacific.edu/engineering/academics/ms-data-science>  
Phone: (209) 946-2151  
Location: San Francisco  
Website: Data Science (<https://engineering.pacific.edu/engineering/academics/ms-data-science/>)

## Degrees Offered

Master of Science in Data Science

## Data Science Program Overview

The MS in Data Science prepares graduates for careers in data analytics and related fields. This is a science (as opposed to business) based program that is focused on developing students' math foundation in statistics and linear algebra, and computer programming to prepare them for coursework in topics like machine learning, time series analysis, customer analytics, and data visualization.

This 32-unit, 4-semester degree culminates in an experiential Capstone Project, in which students work on a real-world analytics problem.

Prerequisite entry requirements include:

- A Bachelors degree
- GPA of 3.0 or above
- Educational qualifications and/or work experience in:
  - Statistics
  - Computer programming (any language, although Python and R are preferred)
  - Basic calculus (derivatives)
- In addition, international students must also have:
  - The US equivalent of a GPA of 3.0 or above
  - TOEFL (or equivalent) English language proficiency. A minimum score of 90 or a score of at least 550 (213 on the computer-based test) is required.
  - Official, course-by-course evaluation of their transcripts with an overall U.S. GPA equivalent from one of the agencies accepted by the University.

## Data Science Program Educational Objectives

The MS in Data Science prepares graduates for careers in predictive modeling, artificial intelligence, and related fields. This is done by developing students' math foundations in statistics and linear algebra, and programming skills in the areas of machine learning, data preparation, data modeling, and a variety of data science / analytic solution areas such as customer analytics, fraud detection and healthcare analytics.

The education that students receive will allow them after graduation to:

- Extract value from data to assist organizations in understanding past performance, predicting future events, and optimizing processes;
- Apply the methods of data wrangling, analytic programming, data mining, quantitative methods, modeling, to prepare very large data sets for analysis;
- Design and develop practical data oriented solutions using modern analytic techniques such as machine learning, time series analysis, and clustering;

- Apply the scientific method to develop and test hypotheses using mathematical and statistical principles;
- Conduct compelling communications through informative visualizations and effective presentation skills.

### Data Analysis

- Analyze various forms of data (e.g. numerical, categorical, textual, objects, etc.) using appropriate mathematical and/or machine learning techniques.

### Data Engineering

- Apply modern programming and data engineering skills, extract data from files, databases, or online resources, and transform it for appropriate analysis.

### Professional Presentation

- Effectively communicate results in a format that is appropriate to the audience, via written, oral, and graphical media.

## Master of Science in Data Science

Students must complete a minimum of 32 units with a Pacific cumulative grade point average of 3.0 to earn the master of science in data science degree.

ANLT 201	Linear Algebra for Data Science	1
ANLT 202	Frequentist Statistics	1
ANLT 212	Analytics Computing for Data Science	1
ANLT 221	Introduction to Machine Learning	1
ANLT 222	Machine Learning for Data Science	1
ANLT 224	Data Wrangling	1
ANLT 232	Introduction to Data Visualization	1
ANLT 233	Dynamic Visualization	1
ANLT 242	Relational Databases	1
ANLT 251	Data Science Socratic Lab	3
ANLT 282	Capstone Project	2-6
ANLT 283	Weekly Hot Topics	1

Select five (5) of the following:

ANLT 203	Bayesian Statistics	1
ANLT 207	Time Series Analysis	1
ANLT 210	Software Methods for Data Science	1
ANLT 214	Data Engineering for Data Science	1
ANLT 223	Advanced Machine Learning	1
ANLT 234	Analytics Storytelling for Data Science	1
ANLT 243	NoSQL Databases	1
ANLT 273	Fraud Detection	1
ANLT 274	Customer Analytics	1
ANLT 272	Healthcare Case Studies	1
ANLT 276	Emphasis Case Studies	1

\* Students will take three semesters of both ANLT 251 or ANLT 283 .

## Course Descriptions

### **ANLT 201. Linear Algebra for Data Science. 1 Unit.**

Linear algebra is the generalized study of vector spaces and transformations in n-dimensions. In this course, students begin by developing an understanding of the concepts and operations of linear algebra, which are frequently employed in the analysis of data. Topics include: formulating and solving linear systems as matrix-vector equations, performing basic computations involving matrix algebra, orthogonal projections, and eigenvalue-eigenvector problems. Students are then exposed to important topics for data science, such as singular value decomposition and principle component analysis. The use of software to perform computations is emphasized. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 202. Frequentist Statistics. 1 Unit.**

A survey of regression, linear models, and experimental design. Topics include simple and multiple linear regression, single- and multi-factor studies, analysis of variance, analysis of covariance, model selection, and diagnostics. This class will focus more on the application of regression methods than the underlying theory through the use of modern statistical programming languages. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 203. Bayesian Statistics. 1 Unit.**

This course introduces Bayesian statistical methods that enable data analysts and scientists to combine information from similar experiments, account for complex spatial, temporal, and other relationships, and also incorporate prior information or expert knowledge into a statistical analysis. This course explains the theory behind Bayesian methods and their practical applications, such as social network analysis, predicting crime risk, or predicting credit fraud. The course emphasizes data analysis through the use of modern analytic programming languages. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 205. Consumer Analytics. 1 Unit.**

This course introduces the techniques used to analyze consumer shopping and buying behavior using transactional data in industries like retail, grocery, e-commerce, and others. Students will learn how to conduct item affinity (market basket) analysis, trip classification analysis, RFM (recency, frequency, monetary) analysis, churn analysis, and others. This class will teach students how to prepare data for these types of analyses, as well as how to use machine learning and statistical methods to build the models. The class is an experiential learning opportunity that utilizes real-world data sets and scenarios. Prerequisite: Graduate status in the Data Science program.

### **ANLT 206. Sentiment Analysis and Opinion Mining. 1 Unit.**

This course introduces the algorithms and methods used to analyze the subjective opinions and sentiments of the author of a free text document such as a tweet, blog post, or article. The class will examine the applications of this type of analysis as well as its benefits and limitations. Sentiment analysis is closely tied to text mining and uses techniques such as natural language processing, text analysis, and computational linguistics for feature extraction and preprocessing of the data. Students will explore the current state of usage of sentiment analysis, as well as future implications and opportunities. Prerequisite: Graduate status in the Data Science program.

### **ANLT 207. Time Series Analysis. 1 Unit.**

This course introduces the theory and application of statistical methods for the analysis of data that have been observed over time. Students will learn techniques for working with time series data and how to account for the correlation that may exist between measurements that are separated by time. The class will concentrate on both univariate and multivariate time series analysis, with a balance between theory and applications. Students will complete a time series analysis project using real-world scenario and data set. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 208. Research Methods for Data Science. 1 Unit.**

Students learn about research design, qualitative and quantitative research, and sources of data. Topics will include a variety of research topics, including such things as data collection procedures, measurement strategies questionnaire design and content analysis, interviewing techniques, literature surveys; information databases, probability testing, and inferential statistics. Students will prepare and present a research proposal (with emphasis on technical writing/presentation principles) as part of the course. Prerequisite: Graduate status in the Data Science program.

### **ANLT 210. Software Methods for Data Science. 1 Unit.**

Students learn the tools, methodology, and etiquette in developing data science applications, tools, and analytical workflows in collaborative environments. Data scientists are at the nexus of software engineering, science, and business. In order to thrive in this world, they must work collaboratively across these fields and skill sets, while ensuring that work is accessible and digestible to everyone involved. Moreover, they must ensure their work is production-worthy and extensible. This course teaches all of the elements, both technical and conceptual, to create productive, helpful, and professional data scientists. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 212. Analytics Computing for Data Science. 1 Unit.**

This course introduces computational data analysis using multi-paradigm programming languages. By the end of the course, students will tackle complex data analysis problems. The course emphasizes the use of programming languages for statistical and machine learning analysis, and predictive modeling. Graphical analytics tools will also be used. The course will also cover the various packages for accessing data that come with the various languages, manipulating and preparing data for analysis, conducting statistical and machine learning analyses, and graphically plotting and visualizing data and analytical results. The course emphasizes hands-on data and analysis using a variety of real-world data sets and analytical objectives. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

### **ANLT 214. Data Engineering for Data Science. 1 Unit.**

This course introduces students to data warehousing architectures, big data processing pipelines, and in-memory analytic techniques. Students will learn how to design systems to manage large volumes of multidimensional data. Currently, this includes the map-reduce paradigm, distributed file systems (HDFS), The Spark distributed computing platform, and how to sign up cloud computing resources (AWS EC2). Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 216. Legal Analytics for Data Science. 1 Unit.**

This course introduces students to how the law applies to the practice of data science. This course will expose students to: the ways in which data science assists with the practice of law, legal compliance and regulations that affect how data science tasks can be conducted, and the diverse ways in which the law affects the data scientist in his/her capacity as a practicing professional. Pre-requisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 221. Introduction to Machine Learning. 1 Unit.**

This course introduces the concepts of machine learning at the first-semester graduate level. The course begins with a brief review of linear algebra with applications to data manipulation. Next, linear and logistic regression, SVMs, classification, and clustering are reviewed. Data wrangling methods and concepts such as imputation, transformation, and dimensional reduction are discussed, followed by an introduction to model validation. The last third of the course introduces modern machine learning models and concepts: neural networks, deep learning, decision trees, and natural language processing. Prerequisites: Graduate standing or permission of the MS Data Science program director. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 222. Machine Learning for Data Science. 1 Unit.**

Machine learning is the artificial intelligence discipline for uncovering patterns and relationships contained in large data sets. Students will be exposed to the supervised learning methods such as neural networks and decision trees. Practical application of these techniques will be tools like R and Python. Students will also learn: proper techniques for developing, training, and cross-validating predictive models; bias versus variance; and will explore the practical usage of these techniques in business and scientific environments. Students will also be introduced to unsupervised learning – the class of machine learning for uncovering patterns and relationships in data without labeling the data or establishing a preconceived set of classes or results. Students will learn through hands-on programming projects. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 223. Advanced Machine Learning. 1 Unit.**

This course builds on the fundamentals introduced in ANLT 222 Machine Learning, by examining more machine algorithms and neural network topologies, and studying their respective applications. The course includes an overview of the TensorFlow language, Decision Tree methods, and an introduction to Natural Language Processing (NLP). Prerequisite: ANLT 222 Machine Learning, Graduate status in the MS Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 224. Data Wrangling. 1 Unit.**

This course will teach you how to retrieve data from disparate sources, combine it into a unified format, and prepare it for effective analysis. This aspect of data science is often estimated to be upwards of 80% of the effort in a typical analytics process. Students will learn how to read data from a variety of common storage formats, evaluate its quality, and learn various techniques for data cleansing. Students will also learn how to select appropriate features for analysis, transform them into more usable formats, and engineer new features into more powerful predictors. This class will also teach students how to split the data set into training and validation data for more effective analytical modeling. Prerequisite: Graduate status in the Data Science program.

**ANLT 232. Introduction to Data Visualization. 1 Unit.**

This course introduces tools and methods for visualizing data and communicating information clearly through graphical means. The class covers various data visualizations and how to select the most effective one depending on the nature of the data. Students will practice using the data visualization methodology by walking through a case study with the instructor and then practicing the steps on their own. Students will work with modern analytic graphics packages, and will be introduced to open source libraries, and to commercial visualization products. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 233. Dynamic Visualization. 1 Unit.**

This course introduces advanced visualization techniques for developing dynamic, interactive, and animated data visualization. Students will learn a variety of techniques for the visualization of complicated data sets. These techniques are valuable for visualizing genomic data, social or other complex networks, healthcare data, business dynamics changing over time, weather and scientific data, and others. Often the visual presentation of data is enhanced when it is made interactive and dynamic, allowing users to “move through” the data and manipulate the data graphically for exploratory analysis. This presentation often involves web application development, and students will be exposed to these rudiments as well as tools that enable faster development of data visualization. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 234. Analytics Storytelling for Data Science. 1 Unit.**

This course builds upon ANLT 232. It will dive into how visualizations should be presented differently when presenting to lay people, business executives, and a technical group. It will also consider visualizations meant for exploratory analysis versus persuasive argument versus survey, or “30,000 foot” analysis. Working alone and in teams, students will create visualizations using their own findings and using provided case studies. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 242. Relational Databases. 1 Unit.**

This course introduces relational database management systems (RDBMS) and the structured query language (SQL) for manipulating data stored therein. The class is focused on the applied use of SQL by data scientists to extract, manipulate and prepare data for analysis. Although this class is not a database design class, students will be exposed to entity-relationship (ER) models and the benefits of third normal form (3NF) data modeling. The class employs hands-on experiential learning utilizing the modern relational database querying languages and graphical development environments. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 243. NoSQL Databases. 1 Unit.**

This course will examine different non-relational (NoSQL) database paradigms, such as Key-Value, Document, Column-family, and Graph databases. Students will learn about advantages and disadvantages of the different approaches. The class will include hands-on experience with a representative sample of NoSQL databases. Computing developments that spurred the existence of NoSQL databases, such as big data, distributed and cloud computing will also be discussed. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 251. Data Science Socratic Lab. 3 Units.**

This course is a co-requisite of the asynchronous online courses in the Masters in Data Science program (MSDS). In the Socratic Lab, students will engage with all material being studied by their cohort in the current term, both collectively, individually, and in teams. Work includes further applications and investigations of the online curriculum through projects, discussion, presentations, written reports, and programming. The Socratic Lab allows an integrative approach to reinforcing and expanding the topics students are studying elsewhere in the MSDS program. Prerequisite: Graduate status in the MS Data Science program.

**ANLT 272. Healthcare Case Studies. 1 Unit.**

This course ties together the statistical, computational analytics, and database concepts in a series of case studies in the Healthcare sector. Students will examine four separate case studies of the use of data analytics in healthcare. Students will work in teams to dissect these case studies and evaluate the business opportunity, the analysis methodology, the raw data, the feature engineering and data preparation, and the analytical outcomes. Students will present their evaluation and make recommendations for improvements in the analysis and related opportunities. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 273. Fraud Detection. 1 Unit.**

This course introduces the use of analytics to detect fraud in a variety of contexts. This class shows how to use machine learning techniques to detect fraudulent patterns in historical data, and how to predict future occurrences of fraud. Students will learn how to use supervised learning, unsupervised learning, and social network learning for these types of analyses. Students will be introduced to these techniques in the domains of credit card fraud, healthcare fraud, insurance fraud, employee fraud, telecommunications fraud, web click fraud, and others. The course is experiential and will apply concepts taught in prior data wrangling and machine learning courses using real-world data sets and fraud scenarios. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 274. Customer Analytics. 1 Unit.**

This course introduces the techniques used to analyze consumer shopping and buying behavior using transactional data in industries like retail, grocery, e-commerce, and others. Students will learn how to conduct item affinity (market basket) analysis, trip classification analysis, recommender systems, RFM (recency, frequency, monetary) analysis, churn analysis, and others. This class will teach students how to prepare data for these types of analyses, as well as how to use machine learning and statistical methods to build the models. The class is an experiential learning opportunity that utilizes real-world data sets and scenarios. Prerequisite: Graduate status in the Data Science program. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 275. Text Mining. 1 Unit.**

This course introduces the essential elements of text mining, or the extension of standard predictive methods to unstructured text. The class will explore the use of text mining in domains such as digital security, bioinformatics, law, marketing, and social media. Students will be exposed to information retrieval, lexical analysis, pattern recognition, meta-data tagging, and natural language processing (NLP). A large portion of this class will be devoted to the data preparation and wrangling methods needed to transform unstructured text into a suitable structure for analysis. Prerequisite: Graduate status in the Data Science program.

**ANLT 276. Emphasis Case Studies. 1 Unit.**

This course provides a real-world learning opportunity that ties together the concepts and practice of data science through a series of case studies in the finance, manufacturing, telecommunications and retail sectors. Students evaluate the business opportunities and challenges, explore, wrangle, and prepare the raw data, compare, select, implement, and validate statistical and machine learning models. Students present their evaluations and make recommendations for improvements. Corequisite: ANLT 251 Data Science Socratic Lab.

**ANLT 282. Capstone Project. 2-6 Units.**

This course is a culmination of all modules in the MS Data Science program. It provides an experiential learning opportunity that connects all of the materials covered in the MS Analytics program. Students will be formed into teams and assigned to an industry sponsored project. Capstone projects will be agreed in advance with sponsoring companies and will represent real-world business issues that are amenable to an analytic approach. These projects will be conducted in close oversight by the sponsoring company, as well as, a University faculty member and may be conducted on the sponsoring company's premises using their preferred systems and tools, at the sponsoring company's discretion. Prerequisite: Graduate status in the Data Science program.

**ANLT 283. Weekly Hot Topics. 1 Unit.**

This course consists of a set of weekly presentations and discussions around key analytic issues and current case studies. These hot topics will be presented by a combination of guest speakers – industry luminaries in the area of analytics – and University of the Pacific faculty members, including the MS Analytics program director. Many of these topics will be drawn from relevant real-world contemporary analytic stories that reinforce specific elements of the academic content being taught and cannot be predicted in advance. Prerequisite: Graduate status in the Data Science program.

**ANLT 287B. Internship. 1-4 Units.****ANLT 287A. Internship. 1-4 Units.****ANLT 287. Internship. 1-4 Units.****ANLT 293. NLP for Data Science. 1 Unit.****ANLT 297. Graduate Research. 1-6 Units.**