DATA SCIENCE (DSC)

DSC 223 Intro Math of Data Science (3 credits)

This course provides an introduction to basic mathematical and statistical topics needed to understand data science. It includes elementary set theory and counting techniques, discrete probability, descriptive statistics, simple linear regression, basic inferential statistics, and an introduction to linear algebra. This course also provides an introduction to the R statistical software and programming language. *Prerequisites:* MAT 155 or MAT 161

Attributes: Math Beauty, Undergraduate

DSC 225 Data Science for Sports (3 credits)

This course covers player/team performance assessment and comparisons using historical and online data. To explore various sportsrelated data sets, students will learn methods of data cleaning, data visualization, statistical tests, statistical modeling, predictive analysis, and simulation. The focus will be on Tennis, Soccer, Basketball, and Volleyball. Students will learn to code in Python and/or R. *Prerequisites:* (MAT 118 or MAT 128 or MAT 134 or DSC 223 or DSS 210) *Attributes:* Math Beauty, Undergraduate

DSC 325 Essentials of Data Science (3 credits)

This course covers the basic topics in data science. It includes descriptive and inferential statistics, introduction to simple and multiple regression, data visualization, and data cleaning or scrubbing. It also includes an introduction to machine learning topics such as decision trees, k-nearest neighbors, neural networks and clustering. The R software or the Python programming language will be used to visualize and analyze datasets.

Prerequisites: MAT 223 or DSC 223 Attributes: Math Beauty, Undergraduate

DSC 326 Advanced Data Science (3 credits)

This course covers some advanced topics in data science, including recent tools for performing predictive analytics, data visualization, data wrangling, statistical inference, deep machine learning, and software engineering. Various software packages, including TensorFlow, will be used to build predictive models. Whenever appropriate, the mathematical background of predictive models will be covered. Also, one of the main goals is to introduce students to the most important aspects of data science by reinforcing writing efficient code, testing, and debugging while working with large software systems. The course includes several programming projects in Python and/or R. *Prerequisites:* DSC 325 or CSC 346

Attributes: Math Beauty, Undergraduate

DSC 424 Regression and Time Series (3 credits)

The first part of the course covers Generalized Linear Models (GLMs). Topics include exponential family, important link functions, estimations (maximum likelihood estimation, generalized moment matching), diagnostic tests for model validations (graphical methods, chi-square statistics, t and F tests, AIC and BIC, likelihood ratio test), applications of GLMs on real data, prediction and confidence intervals. It also includes penalized regression (ridge and lasso regression, k-nearest neighbors algorithm). The second part of the course covers Time Series Analysis. Topics include an introduction to discrete stochastic processes, random walks, stationary processes, autocorrelation functions, and partial autocorrelation functions, various time series models (exponential smoothing, autoregressive (AR) model, moving average (MA) model, ARMA model), autoregressive conditional heteroskedastic (ARCH) model, generalized ARCH (GARCH) model, variants of GARCH, predictions and their confidence intervals using time series models.

Prerequisites: MAT 223 or (MAT 128 and MAT 155) or (MAT 128 and MAT 161)

Attributes: Math Beauty, Undergraduate

DSC 425 Machine Learning/Data Science (3 credits)

This course provides an introduction to the fields of Machine Learning, Data Science and Predictive Analytics. It includes linear regression, logistic regression, nearest neighbor methods, decision trees, neural networks, clustering, principal components analysis, and resampling methods such as cross-validation and bootstrapping. If time permits, it will also include support vector machines, deep learning methods, and machine learning methods for numerical optimization such as genetic and evolutionary algorithms and swarm intelligence algorithms. The R software will be used to apply statistical and machine learning methods to real data sets. Whenever appropriate, the mathematical background of machine learning methods will be covered. Students will be required to work on a final data analysis project and present their findings in class. This course and DSC 424 (Regression and Time Series) together cover the topics in the SOA (Society of Actuaries) exam in SRM (Statistics for Risk Modeling) and provide an intro to the PA (Predictive Analytics) exam. Also, this course and DSC 424 cover several topics in the CAS (Casualty Actuarial Society) exams in MAS (Modern Actuarial Statistics) I and II. Prerequisites: MAT 223 or DSC 223

Attributes: Math Beauty, Undergraduate