BDA - Big Data Analytics

BDA 200T Elements of Data Science (3 Credit Hours)
This course offers a non-technical introduction to the emerging and interdisciplinary area of data science. Students will be introduced to the development, fundamental tools, and the impact of data science in a wide range of disciplines such as business, the sciences and engineering. Fundamental data visualization techniques and basic concepts of machine learning will be applied through real-life data science projects. Moreover, students will explore the general framework for ethical thinking and practicing data science, the current challenges, the benefits, the potential harms and risks posed by developing data science models and technology.
Prerequisites: MATH 102M or MATH 103M

BDA 401/501 Programming Languages for Data Science (3 Credit Hours)
An introductory course on programming languages and tools which are relevant to data analytics. Each language or tool is introduced as a separate module and incorporates applications in mathematics and statistics. Examples of included programming languages and tools are MATLAB, Python, R and SAS. Additional languages and tools may be covered based on current trends in data analytics. Students will complete hands-on programming assignments throughout the course.
Prerequisites: MATH 312, MATH 316 and STAT 330 or STAT 331

BDA 411/511 Introduction to Machine Learning (3 Credit Hours)
An introductory course on machine learning. Machine Learning is the science of discovering pattern and structure and making predictions in data sets. It lies at the interface of mathematics, statistics and computer science. The course gives an elementary summary of modern machine learning tools. Topics include regression, classification, regularization, resampling methods, and unsupervised learning. Students enrolled are expected to have some ability to write computer programs, some knowledge of probability, statistics and linear algebra.
Prerequisites: MATH 312, MATH 316, and STAT 330 or STAT 331

BDA 431/531 Modern Statistical Methods for Big Data Analytics (3 Credit Hours)
The statistical perspective of data mining is emphasized for majority of the course. Both applied aspects (programming, problem solving, and data analysis) and theoretical concepts (learning, understanding, and evaluating methodologies) of data mining will be covered. Topics include Regularization and Kernel Smoothing Methods, Tree-based Methods, Neural Networks and optional topics such as deep learning.
Prerequisites: BDA 411 and STAT 405

BDA 432/532 Introduction to Optimization in Data Science (3 Credit Hours)
Topics considered include the solution of non-smooth optimization problems arising in data science, including unconstrained and constrained optimization problems, Lagrange multiplier methods, inequality constraints, Kuhn-Tucker conditions, and applications. Also considered are linear and nonlinear inverse problems, regularization of ill-posed problem including singular value decomposition, and Tikhonov regularization methods and sparse regularization methods, inverse eigenvalue problems and applications such as compressed sensing, image reconstruction and machine learning.
Prerequisites: MATH 307, MATH 312 and MATH 316

BDA 450 Senior Project in Big Data Analytics I (3 Credit Hours)
This course introduces students to practical applications of big data analytics. Lecture topics include an overview of the various topics in business, engineering, and government currently using big data analytics. Students will choose a project involving a real world application to explore techniques learned during other course work. Course involves written and oral presentations for students to improve communication and teamwork skills.
Prerequisites: A grade of C or better in STAT 331 and STAT 405
Pre- or corequisite: BDA 431

BDA 451 Senior Project in Big Data Analytics II (3 Credit Hours)
This course allows the student to pursue an in-depth exploration of a project initiated in BDA 450. The course involves written and oral presentations for students to improve communication and teamwork skills.
Prerequisites: BDA 450 and permission of instructor

BDA 501 Programming Languages for Data Science (3 Credit Hours)
An introductory course on programming languages and tools which are relevant to data analytics. Each language or tool is introduced as a separate module and incorporates applications in mathematics and statistics. Examples of included programming languages and tools are MATLAB, Python, R and SAS. Additional languages and tools may be covered based on current trends in data analytics. Students will complete hands-on programming assignments throughout the course.
Prerequisites: MATH 312, MATH 316 and STAT 330 or STAT 331

BDA 511 Introduction to Machine Learning (3 Credit Hours)
An introductory course on machine learning. Machine Learning is the science of discovering pattern and structure and making predictions in data sets. It lies at the interface of mathematics, statistics and computer science. The course gives an elementary summary of modern machine learning tools. Topics include regression, classification, regularization, resampling methods, and unsupervised learning. Students enrolled are expected to have some ability to write computer programs, some knowledge of probability, statistics and linear algebra.
Prerequisites: MATH 312, MATH 316, and STAT 330 or STAT 331

BDA 531 Modern Statistical Methods for Big Data Analytics (3 Credit Hours)
The statistical perspective of data mining is emphasized for majority of the course. Both applied aspects (programming, problem solving, and data analysis) and theoretical concepts (learning, understanding, and evaluating methodologies) of data mining will be covered. Topics include Regularization and Kernel Smoothing Methods, Tree-based Methods, Neural Networks and optional topics such as deep learning.
Prerequisites: BDA 511 and STAT 505 or permission of the instructor

BDA 532 Introduction to Optimization in Data Science (3 Credit Hours)
Topics considered include the solution of non-smooth optimization problems arising in data science, including unconstrained and constrained optimization problems, Lagrange multiplier methods, inequality constraints, Kuhn-Tucker conditions, and applications. Also considered are linear and nonlinear inverse problems, regularization of ill-posed problem including singular value decomposition, and Tikhonov regularization methods and sparse regularization methods, inverse eigenvalue problems and applications such as compressed sensing, image reconstruction and machine learning.
Prerequisites: MATH 307, MATH 312 and MATH 316

BDA 611 Mathematical Foundations of Machine Learning (3 Credit Hours)
This course will introduce mathematical foundations of machine learning theory and algorithms. Topics include statistical learning theory, kernel methods and generative models. Some modern machine learning methods such as dictionary learning, deep learning, online learning, and reinforcement learning may also be included, time permitting. Students enrolled are expected to have some knowledge of probability, linear algebra, optimization, and analysis.
Prerequisites: BDA 511, MATH 518 and STAT 330 or 331

BDA 620 Large-Scale Optimization (3 Credit Hours)
This course will introduce optimization methods for large-scale problems by exploiting special structures including convexity and sparsity. Topics include introduction to convexity, gradient-related methods, dual methods, sparse optimization methods and nonconvex optimization methods. Students enrolled are expected to have some knowledge of linear algebra, optimization, probability, and analysis.
Prerequisites: MATH 518 and STAT 330 or 331

BDA 632 Computational Data Analytics Project (3 Credit Hours)
Under the guidance of a faculty member in the Department of Mathematics and Statistics, the student will undertake a significant computational data analysis problem. A written report and/or public presentation of results will be required.
Prerequisites: Permission of graduate program director

BDA 640 Genomic Data Science (3 Credit Hours)
Introductory discussion on central dogma of molecular biology, concepts of transcription, translation, gene regulation, and the need for high throughput methods. Other topics covered are Introduction to R and Bioconductor, Advanced microarray data analysis, NGS data analysis using edgeR in Bioconductor, Network Biology, sequence, pathway informatics, SNPs, GWAS, informatics for genome variants.
Prerequisites: BDA 511, BDA 531, and STAT 505 or permission of the instructor

BDA 697 Topics in Big Data Science (3 Credit Hours)
Advanced study of selected topics.

BDA 721 High-Dimensional Statistics (3 Credit Hours)
Techniques for obtaining basic tail bounds and concentration inequalities, uniform laws of large numbers, Rademacher complexity of a set, covering and packing in metric spaces, and metric entropy. Also, high dimensional random matrices described in a non-asymptotic framework, with a focus on the estimation of sparse and structured covariance matrix, are studied. The sparse linear regression models and the principal component analysis in the unstructured and sparse setting will be covered.
Pre- or corequisite: STAT 727, STAT 728, MATH 616, and MATH 618

BDA 731 Applied Functional Data Analysis (3 Credit Hours)
An introduction to the statistical analysis of sample curves or functions. Topics include smoothing, registration, functional principal component analysis, scalar-on-function regression, and functional response models. All these techniques will be applied using the statistical software R.
Prerequisites: STAT 725 or STAT 825

BDA 745 Transform Methods for Data Science (3 Credit Hours)
Various transform methods from the data domain to coefficients of the data in certain discrete bases are studied. Transforms studied include FFT, DCT, wavelet transforms and framelet transform. Both theory and applications of these transforms are covered.
Prerequisites: MATH 518 and MATH 616

BDA 821 High-Dimensional Statistics (3 Credit Hours)
Techniques for obtaining basic tail bounds and concentration inequalities, uniform laws of large numbers, Rademacher complexity of a set, covering and packing in metric spaces, and metric entropy. Also, high dimensional random matrices described in a non-asymptotic framework, with a focus on the estimation of sparse and structured covariance matrix are studied. The sparse linear regression models and the principal component analysis in the unstructured and sparse setting will be covered.
Prerequisites: STAT 727, STAT 728, MATH 616, and MATH 618

BDA 831 Applied Functional Data Analysis (3 Credit Hours)
An introduction to the statistical analysis of sample curves or functions. Topics include smoothing, registration, functional principal component analysis, scalar-on-function regression, functional response models. All these techniques will be applied using the statistical software R.
Prerequisites: STAT 725 or STAT 825

BDA 845 Transform Methods for Data Science (3 Credit Hours)
Various transform methods from the data domain to coefficients of the data in certain discrete bases are studied. Transforms studied include FFT, DCT, wavelet transforms and framelet transform. Both theory and applications of these transforms are covered.
Prerequisites: MATH 518 and MATH 616