

KHOURY COLLEGE MEANINGFUL MINORS

APPROVED ELECTIVE COURSES FOR DATA SCIENCE

Northeastern University
**Khoury College of
Computer Sciences**

BOUVÉ COLLEGE OF HEALTH SCIENCES

HINF 5101 Introduction to Health Informatics and Health Information Systems

Introduces the history and current status of information systems in healthcare: information architectures, administrative and clinical applications, evidence-based medicine, information retrieval, decision support systems, security and confidentiality, bioinformatics, information system cycles, the electronic health record, key health information systems and standards, and medical devices.

HINF 5102 Data Management in Healthcare

Explores issues of data representation in healthcare systems, including patient and provider identification, audit trails, authentication, and reconciliation. Discusses underlying design of repositories for electronic health records (EHRs) and computerized provider order entry (CPOE) systems. Includes an overview of privacy issues, legislation, regulations, and accreditation standards unique to healthcare.

HINF 5300 Personal Health Interface Design and Development

Explores the design of innovative personal health human-computer interface technologies. Examples include assistive technologies that aid persons with disabilities, consumer wellness promotion applications, patient education and counseling systems, interfaces for reviewing personal health records, and elder care and social network systems that monitor health and support independent living. Offers students an opportunity to work in teams to build a prototype personal health interface system to solve a real problem. Topics include needs assessment and participatory research, iterative user interface design methods for health interface development, computational sensing of health states and behavior, software architectures for iteratively testing prototype personal health interface technologies, human-computer interaction issues related to personal health technology, and technology transfer requirements to support future validation studies of technology.

HINF 5301 Personal Health Technologies: Field Deployment and System Evaluation

Explores the deployment and evaluation of innovative personal health technologies. In this project-based course, students work in teams to deploy and evaluate a prototype personal health technology that has been previously developed by students in HINF 5300. Offers students an opportunity to develop a research plan to measure the effectiveness, usability, and/or feasibility of the technology; recruit study participants; deploy the technology; and analyze the data collected. Also offers students an opportunity to learn about each of these steps and work toward producing a publishable-quality research paper on the technology and results of the efficacy study, as well as to prepare a grant application that extends the technology and research methodology. Additional topics include technology transfer and implications on health policy.

LEARN MORE

W: khoury.northeastern.edu/academics/undergraduate/minors/
E: Khoury-Advising@northeastern.edu

COLLEGE OF ARTS, MEDIA AND DESIGN

ARTG3451 Information Design 1

Introduces basic concepts, methods, and procedures of information design with a focus on mapping information. Students investigate visual systems and information structures such as maps, graphs, charts, and diagrams. Emphasizes the creative process of organizing, visualizing, and communicating data by making complex information easier to understand and use.

ARTG4552 Information Design 2

Builds on concepts from ARTF 2223 and ARTG 3451. Offers students an opportunity to develop strategies for structuring and communicating complex information to increase understanding through dynamic states, which are controlled through the interaction of end users. Explores possibilities offered by interfaces that mediate between a person and information space through research, projects, readings, and discussions.

ARTG5100 Information Design Studio 1—Principles

Explores the theories and practices of information design through studio projects. Investigates visual systems and information structures such as maps, timelines, charts, and diagrams. Emphasizes the creative process of organizing, visualizing, and communicating data by seeking to make complex information easier to understand and use. Requires graduate standing or permission of program coordinator or instructor.

ARTG5110 Information Design History

Investigates the history of visualization practices across disciplines and in relation to technology developments. Critically examines seminal visualizations in social, cultural, and technological contexts by means of discussions and writing activities in a seminar format. Requires graduate standing or permission of program coordinator or instructor.

ARTG5120 Information Design Research Methods

Examines qualitative and quantitative research methods pertinent to information communication systems. Through discussion and writing activities, offers students an opportunity to investigate varied inquiry toward the development of researchable questions, argument formation, and assessment methodologies. Students who do not meet course restrictions may seek permission of instructor or program coordinator.

ARTG5330 Visualization Technologies

Introduces programming languages that allow computational analysis and digital delivery of dynamic information. Examines implications of environmental and personal sensor data sources, mobile collection and analysis of data, real-time networked data sets, and social use of shared data visualization tools. Students who do not meet course restrictions may seek permission of instructor or program coordinator. May be repeated once.

ARTG6100 Information Design Studio 2—Dynamic Mapping and Models

Continues the exploration of data representations in a variety of media. Focuses on interactive and time-based techniques. Emphasizes computational methods of data collection, manipulation, and encoding. Requires graduate standing or permission of program coordinator or instructor.

ARTG6200 Information Design Studio 3—Synthesis

Continues the exploration of theories of information design and visualization through focused projects that are intended to lead to development of a thesis project. Requires graduate standing or permission of program coordinator or instructor.

GSND5110 Game Design and Analysis

Provides theoretical background and foundation for analyzing and designing games. Examines fundamental domains that are necessary to understand what games are and how they affect players, including but not limited to interface design, level design, narrative, learning, and culture. Presents relevant concepts and frameworks from a wide variety of disciplines—psychology, phenomenology, sociology, anthropology, media studies, affect theories, learning theories, and theories of motivation—for each domain. Explains the core elements of game design, introduces students to formal abstract design tools, explores several models of design process and iteration, and offers students an opportunity to practice game design in groups.

GSND6350 Game Analytics

Introduces the topic of game analytics, defined as the process of discovering and communicating patterns in data with a goal of solving problems and developing predictions in user behavior supporting decision management, driving action, and/or improving game products. Covers the fundamental tools, methods, and principles of game analytics, including the knowledge-discovery process, data collection, feature extraction and selection, pattern recognition to aid in prediction and churn analysis, visualization, and reporting. Covers analytics across game forms, notably online games and delivery platforms. Presents analytical tools recommended during development and tools designed for ongoing maintenance of games.

KHOURY COLLEGE OF COMPUTER AND INFORMATION SCIENCES

IA5010 Foundations of Information Assurance

Builds a common cross-disciplinary understanding in the foundations of information assurance. Presents an overview of basic principles and security concepts related to information systems, including workstation security, system security, and communications security. Introduces information security via database technology. Discusses legal infrastructure such as DMCA, Telecommunications Act, wire fraud, and other ethical issues. Covers security methods, controls, procedures, economics of cybercrime, criminal procedure, and forensics. Describes the use of cryptography as a tool, software development processes, and protection.

IA5050 Data Mining in Cyberspace

Focuses on the basics of the technical, legal, social, and ethical issues implicit in commercial data mining ventures. Introduces the key concepts of data science with specific emphasis on applications in information assurance and the ethical treatment of privacy in data mining. Centers on principles and methods covering the process from envisioning the problem to applying data science techniques to deploying the results to improve information assurance. Topics include an introduction of canonical data mining tasks, spam and fraud detection, Sybil attacks, privacy in data mining, privacy in social networks, and management of information assurance in data science projects.

COLLEGE OF ENGINEERING

EECE 4542 Advanced Engineering Algorithms

Covers classical and modern algorithms that efficiently solve hard electrical and computer engineering optimization problems. These problems arise in a wide range of disciplines—including computer-aided design, parallel computing, computer architecture, and compiler design—and are usually NP-complete, making it unlikely that optimal solutions can be found in a reasonable amount of time. Covers the fundamentals of algorithm analysis and complexity theory and then surveys a wide range of combinatorial optimization techniques, including exhaustive algorithms, greedy algorithms, integer and linear programming, branch and bound, simulated annealing, and genetic algorithms. Considers the efficient generation of optimal solutions, the development and evaluation of heuristics, and the computation of tight upper and lower bounds.

IA5200 Security Risk Management and Assessment

Creates the opportunity for competency in the development of information security policies and plans including controls for physical, software, and networks. Discusses different malicious attacks, such as viruses and Trojan horses, detection strategies, countermeasures, damage assessment, and control. Covers information system risk analysis and management, audits, and log files. Uses case studies, site visits, and works with commercial products.

EECE 5639 Computer Vision

Introduces topics such as image formation, segmentation, feature extraction, matching, shape recovery, dynamic scene analysis, and object recognition. Computer vision brings together imaging devices, computers, and sophisticated algorithms to solve problems in industrial inspection, autonomous navigation, human-computer interfaces, medicine, image retrieval from databases, realistic computer graphics rendering, document analysis, and remote sensing. The goal of computer vision is to make useful decisions about real physical objects and scenes based on sensed images. Computer vision is an exciting but disorganized field that builds on very diverse disciplines such as image processing, statistics, pattern recognition, control theory, system identification, physics, geometry, computer graphics, and learning theory.

EECE 5642 Data Visualization

Introduces relevant topics and concepts in visualization, including computer graphics, visual data representation, physical and human vision models, numerical representation of knowledge and concept, animation techniques, pattern analysis, and computational methods. Topics include tools and techniques for practical visualization and elements of related fields, including computer graphics, human perception, computer vision, imaging science, multimedia, human-computer interaction, computational science, and information theory. Covers examples from a variety of scientific, medical, interactive multimedia, and artistic applications. Includes hands-on exercises and projects. Emphasizes modern engineering applications of computer vision, graphics, and pattern classification methodologies for data visualization.

EECE 5644 Introduction to Machine Learning and Pattern Recognition

Studies machine learning, the study and design of algorithms that enable computers/machines to learn from experience/data. Covers a range of algorithms, focusing on the underlying models between each approach. Emphasizes the foundations to prepare students for research in machine learning. Topics include Bayes decision theory, maximum likelihood parameter estimation, model selection, mixture density estimation, support vector machines, neural networks, probabilistic graphics models, and ensemble methods (boosting and bagging). Offers students an opportunity to learn where and how to apply machine learning algorithms and why they work.

COLLEGE OF SCIENCE

BINF 6308 Bioinformatics Computational Methods 1

Offers the first semester of a two-semester sequence on the use of computers in bioinformatics research. Offers students an opportunity to work with current methods and computational algorithms used in contemporary sequence analysis. Teaches practical skills necessary to manage and mine the vast biological information being generated and housed in public databases. Emphasizes the use of Perl as the primary computer language and requires students to learn and understand basic computer logic and syntax, including an introduction to scalars, arrays, hashes, decision statements, loops, subroutines, references, and regular expressions. A focus on fundamental skills, including the command line interface found in the Linux operating system, is designed to prepare students for second-semester applications.

IE 4615 Expert Systems and Neural Networks

Covers the theory and applications of expert systems and neural networks in engineering. Topics include knowledge representation (semantic networks, frames, production rules, and logic systems), problem-solving methods (heuristic search algorithms, forward and backward chaining, constraint handling, truth, and maintenance), approximate reasoning methods (Bayesian, Dempster-Shafer, fuzzy logic, and certainty factors), and expert system shells. Reviews background material on important neural network architectures such as feed-forward neural networks, Kohonen's feature maps, radial basis function networks, and adaptive resonance theory networks. Discusses neural network applications in several areas including group technology; part family formation; manufacturing systems design, process, and machine tool monitoring and diagnosis; system identification and control; and product inspection.

IE 5640 Data Mining for Engineering Applications

Introduces data mining concepts and statistics/machine learning techniques for analyzing and discovering knowledge from large data sets that occur in engineering domains such as manufacturing, healthcare, sustainability, and energy. Topics include data reduction, data exploration, data visualization, concept description, mining association rules, classification, prediction, and clustering. Discusses data mining case studies that are drawn from manufacturing, retail, healthcare, biomedical, telecommunication, and other sectors.

BINF 6309 Bioinformatics Computational Methods 2

Designed to build upon the core topics covered in BINF 6308, i.e., use of the computer as a tool for bioinformatics research. Builds upon the Perl language fundamentals covered during the first semester but requires students to apply these fundamentals to a semester-long project. The project includes protein family analysis, multiple sequence analysis, phylogeny, and protein structure analysis. Additionally, students have an opportunity to learn to build, load, connect, and query custom MySQL databases, parse command line flags, and build Perl objects.

ENVR 2500 Biostatistics

Offers an overview of traditional and modern statistical methods used to analyze biological data using the free and open-source R programming environment. Lectures describe core statistical approaches and discuss their suitability for understanding patterns that arise at different levels of biological organization, from cellular processes to whole ecosystems. Supervised lab sessions offer students an opportunity to develop the R programming skills required to analyze the complex datasets that often emerge when addressing cutting-edge questions in biology. Topics include basic probability and sampling theory, experimental design, null hypothesis significance testing, t-tests and ANOVA, correlation and regression, Monte Carlo simulations, likelihood, generalized linear models, model selection, and information theory.

MATH 2331 Linear Algebra

Uses the Gauss-Jordan elimination algorithm to analyze and find bases for subspaces such as the image and kernel of a linear transformation. Covers the geometry of linear transformations: orthogonality, the Gram-Schmidt process, rotation matrices, and least squares fit. Examines diagonalization and similarity, and the spectral theorem and the singular value decomposition. Is primarily for math and science majors; applications are drawn from many technical fields. Computation is aided by the use of software such as Maple or MATLAB, and graphing calculators.

MATH 3081 Probability and Statistics

Focuses on probability theory. Topics include sample space; conditional probability and independence; discrete and continuous probability distributions for one and for several random variables; expectation; variance; special distributions including binomial, Poisson, and normal distributions; law of large numbers; and central limit theorem. Also introduces basic statistical theory including estimation of parameters, confidence intervals, and hypothesis testing.

COLLEGE OF SOCIAL SCIENCES AND HUMANITIES

ECON 2350 Statistics

Discusses basic probability, descriptive statistics, estimation techniques, statistical hypotheses, sampling, analysis of variance, correlation, and regression analysis in the context of economics. Computer applications are an integral part of the course.

MATH 4581 Statistics and Stochastic Processes

Continues topics introduced in MATH 3081. The first part of the course covers classical procedures of statistics including the t-test, linear regression, and the chi-square test. The second part provides an introduction to stochastic processes with emphasis on Markov chains, random walks, and Brownian motion, with applications to modeling and finance.

PSYC 2320 Statistics in Psychological Research

Offers an integrated lecture/lab one-semester course covering descriptive and inferential statistics with a focus on psychological applications. Includes a lab to provide hands-on experience with important concepts. Covers standard material in undergraduate statistics including distributions, central tendency, variability, z-scores, the normal distributions, correlation, regression, probability, hypothesis testing (using the z, t, F, and Chi-square statistics), and confidence intervals. This course should be taken before the end of the sophomore year.

PHTH 2210 Foundations of Biostatistics

Introduces the fundamental concepts of biostatistics. Offers students an opportunity to learn to apply statistical thinking to practical problems across several health disciplines. Draws examples and readings from clinical and public health literature. Introduces the Stata statistical software package.

ECON 2560 Applied Econometrics

Examines research methods used by practicing economists. Discusses typical problems from applied areas of economics including choice of modeling framework, problems of data collection, review of estimation techniques, interpretation of results, and development of static and dynamic adaptive policy models. A research paper utilizing computer applications is an integral part of the course.

PPUA 5301 Introduction to Computational Statistics

Introduces the fundamental techniques of quantitative data analysis, ranging from foundational skills—such as data description and visualization, probability, and statistics—to the workhorse of data analysis and regression, to more advanced topics—such as machine learning and networks. Emphasizes real-world data and applications using the R statistical computing language. Analyzing and understanding complex data has become an essential component of numerous fields: business and economics, health and medicine, marketing, public policy, computer science, engineering, and many more. Offers students an opportunity to finish the course ready to apply a wide variety of analytic methods to data problems, present their results to nonexperts, and progress to more advanced course work delving into the many topics introduced here.

FINA 4608 Advanced Financial Strategy

Covers strategic financial decision making in dynamic and technology-driven organizations operating in domestic and international settings. Through case studies, discussions with senior financial executives, and student projects, students gain insight into capital investing and financing decisions in the new economy. An analytical paradigm linking business strategy, financial management, and valuation is utilized to explore financial decision making throughout the life cycle of companies, intended to optimize shareholder value creation. Topics include fundamental financial analysis, capital budgeting under conditions of high risk and uncertainty, startup financing, creative financing, mega-mergers, risk management, and valuation.

D'AMORE-MCKIM SCHOOL OF BUSINESS

MISM 3305 Information Resource Management

Examines how information technology is used to support the functional areas of business (finance, accounting, marketing, manufacturing, and human resource management) to achieve business results (creating new products and services, redesigning business operations, and altering relations with customers and suppliers to achieve competitive advantage). Offers students an opportunity to understand the business issues involved in investing in new technologies.

MISM 3403 Data Management in the Enterprise

Offers students an introduction to and overview of the methodological frameworks and tool sets for the design, development, and implementation of data-management solutions. Today, almost no aspect of business operates without a strong reliance on the flow of information. Even small enterprises track huge volumes of data, from sales transactions and supply chain activities to Web site traffic. Knowledge workers and managers at all levels within the organization require an understanding of data management, database design and operations, and associated decision- support and data-analysis tools and systems to complete even day-to-day tasks. Offers students an opportunity to work hands-on, applying these methods and tools to solve actual business problems.

MKTG 3401 Marketing Research

Focuses on the marketing research process and the analysis of data using statistical software. Helps students develop an understanding of consumer attitudes and behavior processes as the basis of the design of marketing problems. Topics include problem definition, research design, sampling, attitude measurement, questionnaire design, data collection, and data analysis. Students are expected to work on group projects. The course requires no previous computer experience.

MKTG 3501 Marketing Analytics

Studies the importance of using an analytical approach to support marketing decision making in organizations and offers students an opportunity to learn how to implement such an approach in practice. Focuses on data science in marketing: identifying and acquiring the right data for addressing different marketing challenges, building skills necessary for conducting relevant quantitative analyses, and guiding how to use obtained insights to make better marketing decisions. Topics may include product innovation, market identification and segmentation, customer valuation, media attribution models, and assessment of digital and social media. Students are expected to apply statistical concepts and use relevant software packages for analyzing marketing datasets.