

DATA SCIENCE

DSCI403. INTRODUCTION TO DATA SCIENCE. 3.0 Semester Hrs.

(I, II) This course will teach students the core skills needed for gathering, cleaning, organizing, analyzing, interpreting, and visualizing data. Students will learn basic SQL for working with databases, basic Python programming for data manipulation, and the use and application of statistical and machine learning toolkits for data analysis. The course will be primarily focused on applications, with an emphasis on working with real (non-synthetic) datasets. Prerequisite: CSCI101 or CSCI102 or CSCI261 or CSCI200.

DSCI470. INTRODUCTION TO MACHINE LEARNING. 3.0 Semester Hrs.

(I) The goal of machine learning is to build computer systems that improve automatically with experience, which has been successfully applied to a variety of application areas, including, for example, gene discovery, financial forecasting, and credit card fraud detection. This introductory course will study both the theoretical properties of machine learning algorithms and their practical applications. Students will have an opportunity to experiment with machine learning techniques and apply them to a selected problem in the context of term projects. Prerequisite: CSCI101 or CSCI 102 or CSCI261 or CSCI200; MATH201, MATH332.

DSCI503. ADVANCED DATA SCIENCE. 3.0 Semester Hrs.

(I, II) This course will teach students the core skills needed for gathering, cleaning, organizing, analyzing, interpreting, and visualizing data. Students will use the python programming language and related toolkits for data manipulation and the use and application of statistical and machine learning for data analysis. The course will be primarily focused on applications, with an emphasis on working with real (non-synthetic) datasets. Students will propose and design a semester project using a dataset from their domain of interest, leveraging the concepts and skills acquired from this course (e.g., data analysis, ethical considerations, evaluation and synthesis of results, storytelling and visualization). Prerequisite: CSCI220 with a grade of C- or higher or CSCI262 with a grade of C- or higher, MATH201 or MATH334 OR Graduate level standing and at least CSCI128 or equivalent.

Course Learning Outcomes

- Acquire, clean, and organize structured and unstructured data from a variety of sources, including raw data files, online repositories, and through the use of web scraping and various APIs.
- Utilize toolkits and exploration to preprocess small, medium, and large datasets for input to statistical and machine learning algorithms, including methods of feature extraction, outlier removal, and dimensionality reduction.
- Apply statistical and machine learning toolkits to small, medium, and large datasets, including applications of regression, classification, clustering, and a brief introduction to neural networks.
- Conduct analysis of results and evaluate the predictive power of various statistical and machine learning techniques.
- Develop storytelling and visualization skills to inform (exploratory) or persuade (explanatory) a specific audience using data.
- Recognize and address the ethical issues arising from data collection and statistical and machine learning.
- Design, propose, and present a semester project using a dataset from their domain of interest leveraging the concepts and skills from this course.

DSCI530. STATISTICAL METHODS I. 3.0 Semester Hrs.

Introduction to probability, random variables, and discrete and continuous probability models. Elementary simulation, data summarization and analysis using the R Data Analysis Environment. Confidence intervals and hypothesis testing for means and variances. Chi square tests. Distribution-free techniques and regression analysis. Students are expected to have knowledge of probability covered in MATH334 or an equivalent course. Prerequisite: MATH334 or equivalent.

DSCI560. INTRODUCTION TO KEY STATISTICAL LEARNING METHODS I. 3.0 Semester Hrs.

Part one of a two-course series introducing statistical learning methods with a focus on conceptual understanding and practical applications. Methods covered will include Introduction to Statistical Learning, Linear Regression, Classification, Resampling Methods, Basis Expansions, Regularization, Model Assessment and Selection. Prerequisite: DSCI530 or MATH530.

DSCI561. INTRODUCTION TO KEY STATISTICAL LEARNING METHODS II. 3.0 Semester Hrs.

Equivalent with MATH561,
Part two of a two course series introducing statistical learning methods with a focus on conceptual understanding and practical applications. Methods covered will include Non-linear Models, Tree-based Methods, Support Vector Machines, Neural Networks, Unsupervised Learning. Prerequisite: DSCI560 or MATH560.

DSCI570. INTRODUCTION TO MACHINE LEARNING. 3.0 Semester Hrs.

(I, II) The goal of machine learning is to build computer systems that improve automatically with experience, which has been successfully applied to a variety of application areas, including, for example, gene discovery, financial forecasting, and credit card fraud detection. This introductory course will study both the theoretical properties of machine learning algorithms and their practical applications. Students will have an opportunity to experiment with machine learning techniques and apply them to a selected problem in the context of term projects. Graduate students must complete a more challenging project that utilizes complex machine learning algorithms, requiring a deeper understanding of machine learning approaches and critical thinking. Prerequisite: CSCI101 or CSCI102 or CSCI128, MATH201 or MATH334, MATH332 OR Graduate level standing and at least CSCI128 or equivalent.

Course Learning Outcomes

- Apply supervised, unsupervised, reinforcement machine learning models and deep learning models to solve problems in areas such as prediction, recognition and classification.
- Explore and develop with various tools, techniques and libraries in Python for data processing, feature extraction, visualization, validation and evaluation.
- Create data visualization tools, techniques, and libraries in Python to visualize high dimensional or complex data for stakeholders.
- Determine ethical implications through interpretability of big data and results from the application of various machine learning models.
- Design and develop a machine learning product that solves their chosen real-world challenge.
- Create a video presentation that succinctly outlines the problem, solutions, conclusions, and lessons learned regarding product development for the stakeholders.

DSCI575. MACHINE LEARNING. 3.0 Semester Hrs.

The goal of machine learning research is to build computer systems that learn from experience and that adapt to their environments.

Machine learning systems do not have to be programmed by humans to solve a problem; instead, they essentially program themselves based on examples of how they should behave, or based on trial and error experience trying to solve the problem. This course will focus on the methods that have proven valuable and successful in practical applications. The course will also contrast the various methods, with the aim of explaining the situations in which each is most appropriate.