DATA ANALYTICS - DAT

DAT 111 Introduction to Reporting and Analysis   3 Credits
An introduction to the methods and tools for reporting quantitative data for decision support in a wide range of fields. This course is meant as an introductory course in the Data Science program, and for students in other disciplines preparing for decision support roles in a range of commercial, educational or research roles. Both the general theories and approaches to the presentation of data for decision support in tabular and graphic forms, and practical technical methods will be covered in the course. Most of the course time will be spent using Excel for these tasks, but Tableau and/or PowerBI as well as some basic SQL queries will also be covered. Whenever possible, "real-world" data drawn from a wide range of fields and disciplines will be used to illustrate problems and approaches to reporting of data.
Fulfills College Core: Field 7 (Mathematical Sciences)
Offered: every spring.

DAT 211 Advanced Statistics with R   3 Credits
This course is designed to introduce students to the programming language R. We will begin by talking about the benefits of R from a practical to an ethical level. Students will learn to install R and load packages. Students will then identify a data set they want to work with over the semester and preregister their project rationale, hypotheses, and analytic plan with OSF. Students will spend the majority of their time learning to execute their analytic plan in R. Students will present their project at Ignatian Scholarship Day. After their ISD presentation, students will archive their materials on OSF and update their preregistration to reflect any modifications made to the plan as they conducted their research, changes they would make if they were going to do the project again, and future analyses they would like to conduct with the data set.
Offered: once a year.

DAT 411 Econometrics   3 Credits
Econometrics is the science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena. Econometric modeling is an important research tool in Economics, Finance, and many other academic disciplines. The goal of this course is to provide you with a basic understanding of Econometric theory and practice. We will focus on model specification, estimation, and testing, using a \hands on" approach. Throughout the course, we will use EXCEL, R, and SAS. We will cover most of Chapters 1-10 of the textbook, followed by some selected special topics as time permits. You should read through each chapter as we cover it. Special emphasis will be placed on conceptual understanding and application of econometric methods. For those who are interested in more involved discussions of the theoretical framework and/or the statistical or mathematical derivation behind any of the ideas discussed in class, feel free to meet with me outside of class.
Prerequisite: MAT 111 and CSC 111 or DAT 211 or MAT 131 or MAT 141 or ECO 255 or ECO 256.
Offered: once a year.

DAT 412 Machine Learning   3 Credits
A foundational development of the core ideas and concepts in machine learning, with emphasis on the statistical foundations of machine learning but also applied work in Python, or a comparable language. Topics covered will include feature engineering and basis sets, gradient descent model fitting, kernel methods, Model selection methods, bootstrapping and other permutation methods, model inference and averaging, tree based methods with boosting and bagging, neural nets and deep learning and graph based methods.
Prerequisite: MAT 219 and CSC 112.
Offered: once a year.

DAT 417 Machine Learning for Natural Language Processing   3 Credits
This course is on constructing, training and using Machine Learning tools (neural networks) for Natural Language Processing, covering the fundamentals of operation of ChatGPT and other tools for generative language applications, translation, theme detection, text summarize, question answering and a range of other applications. This is a programming driven course, in which students will construct and evaluate a number of machine learning applications. Students will construct NLP processing models (neural networks) using the Pytorch and/or TensorFlow frameworks within the python programming language using the Jupyter notebook system. The course will also cover text encoding, tokenization, embedded and other reduced space representations, string and sentence transformations and related topics. Basic predictive models will be covered in the introduction to PyTorch and TensorFlow. Data storage in the Apache Arrow and HuggingFace datasets systems will also be discussed. Students may need to subscribe to the Google Colab Pro platform at a modest cost if they do not have regular access to a computer with an Nvidia GPU. Cost of the subscription is comparable to that of a typical electronic textbook.
Prerequisite: CSC 112 and CSC 112L.
Offered: once a year.

DAT 499 Independent Study Course in Data Science   1-3 Credits
Study and work with a faculty supervisor. Project to be determined by faculty agreement. Independent studies require an application and approval by the associate dean.
Prerequisite: DAT 211.
Offered: every fall & spring.

DAT 500 Interactive Graphical Case Studies in Big Data   1 Credit
Students will be introduced to Data Analytics via the study of a variety of case studies of published studies, or successful commercial applications of methods. Students will also learn to replicate the graphical presentations used in these studies, and develop alternative visual representations of the data used in the studies. The R statistical language will be used, as students learn how to produce publication grade graphics that can be used throughout other courses and in their career.
Offered: every summer.

DAT 511 Data Stewardship: Preparation, Exploration and Handling of Big Data   3 Credits
Data stewardship refers to the process of managing collections of data in an ethical and effective manner, so that business objectives can be achieved efficiently while respecting the rights of individuals. This course will thus cover the substantial ethical issues related to Big Data, but will also address many technical issues related to working with large data sets. Establishing and maintaining quality data poses surprisingly large challenges and can be very time consuming, so that knowledge of effective data cleaning is a key capability for Data Analytics. Students will learn how to download, clean, and prepare data for future analysis, and document the process, as well as understanding how seemingly harmless actions can pose threats to the information security of others.
Prerequisite: CSC 511 or CSC 111.
Offered: every fall.

DAT 512 Statistical Approaches to Big Data   3 Credits
This course is a Core course in the Data Analytics program. It starts with a brief review of univariate statistics and then covers selected topics usually taught in courses in multivariate statistical analysis and regression analysis. It is assumed that every student in this course has completed at least one college-level statistics course. The theoretical knowledge and analytical skills gained in this course are an essential component of the Data Analytics program.
Prerequisite: BAN 609 or equivalent, CSC 512 or equivalent, & MAT 500 or equivalent.
Offered: every spring.

DAT 514 Data Mining and Machine Learning   3 Credits
This course is a Core course in the Data Analytics program. It starts with a brief introduction to Data Mining and Statistical Learning, includes a brief summary of relevant methods covered in a much greater detail in other courses in this program, such as Data Stewardship and Statistical approaches to Big Data, and then covers a number of methods essential in the modern Data Mining and Statistical Learning.
Prerequisites: MAT 500, CSC 511, and CSC 512 or equivalents.
Offered: every spring.

DAT 515 Visualization and Presentation of Advanced Analytics   3 Credits
Students will develop the ability to present complex results from Data Analytics to a range of audiences. The course will cover both real time interactive displays and tools, such as graphic user interface and dashboard design, as well as written, oral and graphical communication of analytic results. Students will complete a range of projects in each of these areas.
Prerequisites: DAT 511 & DAT 521 (courses may be taken concurrently) and the ability to program in Python.
Offered: every spring.

DAT 517 Machine Learning for Natural Language Processing   3 Credits
This course is on constructing, training and using Machine Learning tools (neural networks) for Natural Language Processing, covering the fundamentals of operation of ChatGPT and other tools for generative language applications, translation, theme detection, text summarize, question answering and a range of other applications. This is a programming driven course, in which students will construct and evaluate a number of machine learning applications. Students will construct NLP processing models (neural networks) using the Pytorch and/or TensorFlow frameworks within the python programming language using the Jupyter notebook system. The course will also cover text encoding, tokenization, embedded and other reduced space representations, string and sentence transformations and related topics. Basic predictive models will be covered in the introduction to PyTorch and TensorFlow. Data storage in the Apache Arrow and HuggingFace datasets systems will also be discussed. Students may need to subscribe to the Google Colab Pro platform at a modest cost if they do not have regular access to a computer with an Nvidia GPU. Cost of the subscription is comparable to that of a typical electronic textbook.
Prerequisite: CSC 512 and CSC 512L.
Offered: once a year.

DAT 521 Applied Integrative Projects in Data Analytics I   3 Credits
In this course, students would learn SAS. Since the focus is on hands-on, all lectures would be conducted in a computer lab. Students learn how to input various types of data into SAS, such as text, csv, binary and sas7bdat. How to clean data is an important skill students are expected to master. Students learn how to deal with missing variables and run basic sample statistics such as mean, standard deviation, minimum and maximum. Many visualization techniques would be taught. In addition, students learn how to run some basic statistical functions, such as linear regression. Since this course is a preparation for the next course (DAT 522) titled "Applied Integrative Projects in Data Analytics II", students could start to think about their next big projects.
Offered: every fall.

DAT 522 Applied Integrative Projects in Data Analytics II   3 Credits
This course is supervised internship or project course. Students may chose to apply for a competitive internship position in Data Analytics with a local corporation, government or not-for-profit agency, or may apply to carry out a data analytics project with an employer or on-campus research sponsor.
Prerequisites: DAT 500, DAT 514, DAT 521.
Offered: every fall, spring, & summer.

DAT 555 Seminar on Deep Learning   1 Credit
Deep Learning is a computational and mathematical approach to building "deep" or many layer neural network architectures for solving complex machine learning tasks, such as image processing, audio processing, complex time series, natural language processing and other big data problems. This course would teach students to build and training deep learning models using current state of the art tools.
Prerequisite: CSC 112 or CSC 512 and MAT 500 or MAT 211 or MAT 219.
Offered: occasionally.