DATA ANALYTICS (DAAN)

DAAN 501: Analytics Research and Problem Framing

3 Credits

Students in this course will explore the elements of the research process within quantitative, qualitative, and mixed methods approaches as it applies to research into data analytics and its use. The ethical principles and challenges of research will be covered including human-subject research guidelines and the Institutional Review Board approval process. Students will use these theoretical underpinnings to begin to critically review literature in the analytics domain, determine how research findings are useful in forming their understanding of their work, and place their own research within the context of the extant literature.

Prerequisite: STAT 500

DAAN 545: Data Mining

3 Credits

Practical benefits of data mining will be presented; data warehousing, data cubes, and underlying algorithms used by data mining software.

DAAN 600: Thesis Research

1-15 Credits/Maximum of 15

Thesis Research

Prerequisite: DANN 501

DAAN 822: Data Collection and Cleaning

3 Credits

This course focuses on the tools and techniques required for collecting data and preparing them for further analysis. The presence of incorrect and inconsistent data can significantly distort the results of the analysis often negating the potential benefits of information-driven approaches. As a result a variety of research over the last decades has focused on data cleansing: computational procedures to automatically or semi-automatically identify - and, when possible, correct - errors in large data sets. The goal of this course is to explore and discuss different data collection tools and techniques in addition to learning skills for retrieving data from existing databases. To further enforce data quality and reliability this course will cover techniques for error detection and data cleaning on large databases. Students will learn the available tools and techniques for data collection including automated data collection for databases, retrieving data from available databases, data preparation and cleansing techniques, data quality and reliability and finally learn techniques to identify issues in data collection and how to clean the data.

Prerequisite: STAT 500

DAAN 825: Large-Scale Database and Warehouse

3 Credits

This course provides a broad exploration of current and emerging practices for handling large quantities of data using large-scale database systems. Data is being generated at an exponential rate and handling and analyzing such data needs highly customized tools and processes to handle data-intensive tasks. In particular, this course investigates methods to effectively design, develop, and implement the two dominant types of large-scale databases: data warehouses for dimensional data and NoSQL databases for loosely-structured data. Students will learn to design a wide variety of large database solutions, apply extract-transform-load (ETL) strategies, maintain and evolve large-scale databases, explore the fundamentals of NoSQL systems, and understand the properties of different database technologies against atomicity, consistency, isolation, and durability (ACID) properties.

DAAN 826: LARGE SCALE DATABASES FOR REAL-TIME ANALYTICS

3 Credits

This course provides an exploration of current and emerging big data solutions for handling large quantities of data in real-time. In particular, this course investigates methods to design, develop, and implement several systems used for real-time data analysis and storage such as document databases, column-based databases, queueing systems, and real-time processing systems. Students will learn to design a wide variety of large database solutions, and how to interconnect those systems to create a lambda architecture. Using this platform, students will collect, process, store, and report real-time data.

Prerequisite: DAAN 825

DAAN 846: Network and Predictive Analytics for Socio-Technical Systems

3 Credits

The objective of this course is to provide a foundation in the principles of network and predictive analytics along with hands-on experience with statistical analysis software for studying the interrelatedness of cyber-social and cyber-technical aspects of our society as a whole that have transformed physical communities into virtual communities. Fundamental principles of network and predictive analytics, the importance of studying network structures, and how network structures can facilitate communication, coordination and cooperation will be discussed. Statistical analysis software will be used for analyzing the structure of an organization or a society as whole to detect and capture the dynamic patterns of group membership and structure, and predict threats, attacks, criminal behavior and evolution of criminal networks.

Cross-listed with: INSC 846

DAAN 862: Analytics Programming in Python

3 Credits

This course will explore the development of analytics systems and the application of best practices and established software design principles using the Python programming language and its several toolkits. Students will manipulate, analyze and visualize complex data sets and implement statistical, machine learning, information visualization, text analysis, and social network analysis techniques through popular Python toolkits to gain insight into their data.

DAAN 871: Data Visualization

3 Credits

This course provides a foundation in the principles, concepts, techniques and tools for visualizing large data sets. DAAN 871 Data Visualization (3) The course provides a foundation in the principles, concepts, techniques and tools for visualizing information in large complex data sets. Unlike scientific visualization, which focuses on the presentation of data that has a spatial or physical correspondence, data visualization focuses on mapping complex, abstract information to a physical representation. The development of effective visualization strategies is crucial for not only facilitating an understanding of large complex data sets but also for driving knowledge discovery and the decision making processes in a given domain. In this course, students will learn the key principles involved in data visualization and will explore a wide range of visualization approaches that can be applied for understanding complex data across different data types. Specifically, techniques for visualizing one-dimensional data (e.g., temporal data); two-dimensional data (e.g., geospatial data); multidimensional data (e.g., mapping relational data in n-dimensional space); hierarchies and graphs (e.g., tree structures); networks (e.g., social networks) and text (e.g., mining text and hypertext from Web) will be discussed. Emphasis will be placed on the identification of patterns, trends and differences in visualizations of data from variety of domains (e.g., science, business, engineering, social media, etc.). In addition, students will gain hands-on experience with a variety of visualization tools including: Gephi, ManyEyes, Excel, Science of Science (Sci2), Pajek, Lattix, R, Cfinder, MapEquation, NodeXL, and/or Gapminder.

DAAN 881: Data-Driven Decision Making

3 Credits

Application & interpretation of analytics for real-life decision making. DAAN 881 Data-Driven Decision Making (3) The theory and application of several quantitative decision-making tools will be studied. The usefulness of these tools will be illustrated using projects and case studies throughout the course. Emphasis will be placed on the application of the tools and techniques and the results they generate. Finding patterns in data and appropriately grouping them are essential in the extraction of information in large datasets. This course will use Principal Component Analyses to transform highly correlated sets of data by means of orthogonal transformation. Cluster analysis will be used to properly group data when working with large datasets. When the outcomes involve categorical variables, Logistic regression techniques will be used to estimate the probabilistic values of the output. The decision space will be divided into smaller regions using Regression tree analyses. When factors are too numerous and highly collinear, Partial Least Square Regression methods will be performed.Public access datasets in the healthcare, transportation and finance industries will be used to demonstrate the applications and the limitations of these techniques.

Prerequisite: STAT 500 and DAAN 501

DAAN 888: Design and Implementation of Analytics Systems

3 Credits

Design and implement data science and analytics systems using contemporary tolls and techniques.

Prerequisite: IN SC521 and DAAN 825 and DAAN 881

DAAN 897: Special Topics

1-9 Credits/Maximum of 9

Formal courses given on a topical or special interest subject which may be offered infrequently.