Prerequisite: STAT 500
DAAN 572: Reinforcement Learning
3 Credits
Reinforcement learning, along with supervised learning and unsupervised learning, is one of the three basic types of machine learning. Applications of reinforcement learning span across medical intervention, robotics, game playing, autonomous driving, financial trading, and marketing, among many others. This course will cover the main theory and approaches of reinforcement learning, along with deep learning and common software libraries and packages.

DAAN 600: Thesis Research
1-15 Credits/Maximum of 15
Thesis Research
Prerequisite: DANN 501
DAAN 822: Data Collection and Cleaning
3 Credits
Tools and techniques required for data collection and computational procedures to automatically or semi-automatically identify and eliminate errors in large datasets. DAAN 822 Data Collection and Cleaning (3) 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 correct errors in large datasets. The goal of this course is to explore 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 introduce 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 assessment, and finally learn techniques to identify issues in data collection and how to clean the data.

Prerequisite: STAT 500 and IN SC521
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