

CAP 4770 Intro Data Mining & Machine Intelligence

Credits: 3 credits

Text book, title, author, and year: Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, Ian H. Witten, **Series:** The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann; 3 edition (January 20, 2011), **ISBN-10:** 0123748569, **ISBN-13:** 978-0123748560

Reference materials: Research papers which will be distributed in the class

Specific course information:

Catalog description: This course deals with the principles of data mining. Topics include machine learning methods, knowledge discovery and representation, clustering, classification and prediction models, and social network analytics.

Prerequisites: STA 4821 and COP 3530

Specific goals for the course:

To provide certain technical skills that are important in computer science and engineering applications; to provide knowledge base for data mining principles and machine learning algorithms. To provide hands-on experiences on building classification and clustering methods, and applying methodologies learned in the class for real-world applications.

The goal of this class is for students to gain hands-on experiences on data mining, machine learning, and artificial intelligence methods. Course covers techniques used to collect, analyze, and understand the data from different domains. At the end of the class, students should be able to understand the whole process of data collection and analytics, and carrying out system design for search and mining the data. We will use online web documents (such as Twitter data) as the testbed and practice web mining techniques.

Brief list of topics to be covered:

Data Collection

- Data format
- Data cleansing

Classification and Clustering

- Decision trees
- Neural networks
- Nearest neighbor classification
- K-means clustering

Feature Selection

- Relief-F feature selection
- Information Gains

Social Network Analysis

- Link analysis, PageRank
- Social mining algorithms.