COURSE AND CONTACT INFORMATION

Course: STAT 6289 DATA MINING
Semester: FALL 2016

INSTRUCTOR

Name: Dr. Tatiana (Tanya) V Apanasovich,
Campus Address: Rome Hall, 801 22nd St NW, office 769
Phone: 202-994-9721
E-mail: apanasovich@gwu.edu
Office hours: MON 4:30-6:00

COURSE DESCRIPTION

Introduces basic data mining concepts and techniques for discovering interesting patterns hidden in large-scale data sets, focusing on issues relating to effectiveness and efficiency. Topics covered include data preprocessing, data warehouse, association, classification, clustering, and mining specific data types such as time-series, social networks, multimedia, and Web data. Here is an outline of the course material.

1. Introduction: What is data mining, What is not data mining, Challenges, Other issues
2. Data: Types of data, Data quality, Data preprocessing
3. Classification: Problem definition, General approach, Decision tree induction, Rule based classifiers, Nearest neighbour classifiers, Bayesian classifiers, Artificial neural networks, Support vector machine, Ensemble methods, Model evaluation
4. Association analysis: Problem definition, Frequent itemset generation, Rule generation, Challenges, Interestingness measures, Generalization of association patterns
5. Cluster analysis: Introduction, Similarity and distance, Characteristics of clustering algorithms, Center based clustering techniques, Hierarchical clustering, Density based clustering, Other clustering techniques, Scalable clustering algorithms, Cluster evaluation
6. Visualization: Introduction, General concepts, Visualization techniques
7. Deep Learning

COURSE PREREQUISITE(S)

You are expected to have a basic knowledge of statistics (STAT 2118 Regression Methods or a similar course that covers analysis of research data through simple and multiple regression and correlation)
We will use R primarily for its libraries in statistics and machine learning rather than as a general purpose programming environment.
This class is less about specific tools and more about concepts and techniques

TEXTS

Data Mining: Concepts and Techniques, Third Edition by Jiawei Han, Micheline Kamber and Jian Pei.

OBJECTIVES

The student will be able to properly formulate data mining problems, understand and utilize fundamental data mining techniques, and evaluate the data mining results.
GRADING

Your final grade will be a weighted average of your homework average (15%), in-class exam (40%), and take home project (45%).

CLASS POLICIES

Homework: There will be 7-9 homework assignments, with greater frequency in the first half of the course. Some assignments will be more analytical, others will deal with data analysis and implementation of procedures in R. Assignments must be completed by students individually, but group discussion is permitted. The due time will be 11:59 pm of the due date (usually class date). You have to use Dropbox to submit the homeworks. No late homework will be accepted, but the lowest score will be dropped.

Midterm Exam: One open-book in-class examination will be given on Nov, 14th. Make-up midterm will be given only in exceptional circumstances (e.g. well-documented medical problems).

Take Home Project: Take Home Project is an INDIVIDUAL project. You will be asked to build a predictive model to be evaluated on a new data set. Your final product will be R (or any other software) routine, well documented and appropriately commented. The deadline is December, 12th.

ACADEMIC INTEGRITY

I personally support the GW Code of Academic Integrity. It states: “Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information.” For the remainder of the code, see: http://www.gwu.edu/~ntegrity/code.html

SUPPORT FOR STUDENTS OUTSIDE THE CLASSROOM

DISABILITY SUPPORT SERVICES (DSS)
Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Marvin Center, Suite 242, to establish eligibility and to coordinate reasonable accommodations. For additional information please refer to: http://gwired.gwu.edu/dss/

UNIVERSITY COUNSELING CENTER (UCC) 202-994-5300
The University Counseling Center (UCC) offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include:
- crisis and emergency mental health consultations
- confidential assessment, counseling services (individual and small group), and referrals
  http://gwired.gwu.edu/counsel/CounselingServices/AcademicSupportServices

SECURITY

In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.