Fall 2020

CS 634-851: Data Mining

Jason Wang

CS 634 Data Mining

Instructor

Name: Dr. Jason Wang
Office: GITC 4211
Phone: (973) 596-3396
Email: wangj@njit.edu

Prerequisite and Required Background

Fluency in a programming language (Python or Java) is required. Substantial coding and programming will be required in this course.

Course Description

Covers the concepts and principles of association rule mining, classification, clustering, text mining, Web mining, time series data mining and graph mining. Hands-on experiences include the design and implementation of (1) an association rule mining tool for transaction databases, (2) a data clustering tool, a text mining tool or two data classification tools. Some of these projects can be implemented in a cloud computing environment such as Amazon web services. Students can choose a project on Big Data mining with Hadoop/MapReduce/Spark, or either CouchDB or MongoDB technologies. In addition, students have an option to implement a data mining tool using Deep Learning libraries such as TensorFlow and Keras.

Course Objectives

To familiarize students with basic data mining principles, modern data mining methods and tools, as well as advanced data mining applications, and to help students find jobs in fields related to data

mining, machine learning, data science, data analytics, data management and Big Data. Specifically, students will be able to

- Explain data mining and machine learning concepts, principles and methods,
- Use a wide range of publicly available data mining and machine learning tools,
- Evaluate the effectiveness and efficiency of these data mining and machine learning tools based on different performance measures,
- Design, develop and implement custom data mining and machine learning algorithms, heuristics, methods, techniques and software tools.

Recommended Course Textbooks (these books are optional, not required)

- Data Mining: Concepts and Techniques, Han et al., Elsevier, 2011, ISBN 978-0-12-381479-1.
- Introduction to Data Mining, Tan et al., Pearson, 2019, ISBN-13: 978-0-13-312890-1.
- Deep Learning, Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press, 2016.

Course Outline and Schedule

This course covers the concepts and techniques of

1. Association Rule Mining

2. Classification, Regression, Prediction and Forecasting (Supervised Machine Learning)

3. Clustering (Unsupervised Machine Learning)

4. Text Mining

5. Web Searching, Mining and Crawling

6. Mediators, Wrappers and Data Warehousing

7. Time Series Data Mining

8. Graph Mining

9. Advanced Data Mining (Deep Learning)

10. New Applications (Data Science Applications)

Course Workload
There will be one midterm project, one final term project, one term paper, and one exam.

Course Grade
Midterm Project -- 25%, Final Term Project -- 30%, Term Paper -- 10%, Exam -- 35%.

Grading Scale
A: 93% and above; B+: 86%-92.9%; B: 78%-85.9%; C+: 70%-77.9%; C: 60%-69.9%; F: Below 60%.

Honor and Policy

- Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an "F" grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified.

- In the exam, each student is required to sign the Honor Code Agreement "On my honor, I pledge that I have not violated the provision of the NJIT Student Honor Code."