CS831 Knowledge Discovery in Databases

Fall 2020

Instructor: Lisa Fan
CW308.17
Tel: 585-4110
Lisa.Fan@uregina.ca

Lectures:
Time: MWF 10:30am – 11:20am
Location: CL313

Office Hours:
W: 1:30pm – 3:00pm  F: 1:30pm – 3:00pm or by appointment

Starting Date: Sept.2, 2020

Ending Date: Dec.8, 2020

Texts
1. Recently published research papers.

Grading

20% --- Assignments
40% --- Final project report, due Dec.3/2020.
10% --- Final project presentation.
15% --- Final examination.
5% --- Class participation
5% --- instructor’s discretion
Calendar Description: Knowledge discovery from databases in the nontrivial extraction of implicit, previously unknown, and potentially useful information from databases. This course focuses on data sources, extraction techniques, efficiency concerns, and measures of novelty and usefulness. Prior to registering for this course, students should have a background in database and information retrieval, and artificial intelligence comparable to the senior undergraduate level.

Tentative Topics

- Introduction and overview
- Data
  - Types of Data
  - Data preprocessing
  - Measures of similarity and dissimilarity
- Classification
  - Decision tree induction
  - Model overfitting
  - Evaluating the performance of a classifier
- Classification: Alternative techniques
  - Different classifiers
  - Class imbalance problem
- Association Analysis
  - Frequent itemset generation
  - Rule generation
  - Advanced concepts: categorical attributes and continuous attributes; Concept hierarchy; sequential patterns; infrequent patterns
- Clustering Analysis
  - K-means
  - Agglomerative Hierarchical clustering
  - DBSCAN
  - Prototype-based
  - Density–based
  - Graph-based
- Recent Developments on Above Topics
- Advanced Techniques of Data Analysis on Health Informatics.
- Time series
- Web Data Mining
  - Web content mining
  - Web structuring mining
  - Web log mining