CMPT 459 - D100 Special Topics Database Systs

**Instructor(s):**  
Jian Pei

**Calendar Objective/Description:**  
Special Topics Database Systs

**Instructor's Objectives:**

What should you do when you are facing (a huge amount of) data from applications? How can you become a data scientist? This course introduces an important research and development frontier: knowledge discovery in databases (KDD), also known as data mining (DM). This course emphasizes data-driven thinking and focuses on big data, fundamental methods and killer applications.

Students taking this course are expected to (1) have taken an entry-level college statistics course; (2) have taken an algorithm course; and (3) be able to program fluently in C/C++/C#, Java, Python or R.

After taking the course, a student should understand the essential principles of data science, learn a handful of data mining techniques and have a good command of data mining tools.

**Prerequisites:**

see go.sfu.ca

**Topics:**

- Introduction
- Business intelligence, data warehousing, data lakes, and enterprise data infrastructure
- Finding useful patterns
- Predictive analytics
- Clustering analysis
- Anomaly and fault detection
- Working with data as a data scientist

**Grading:**

Grading scheme will be announced in the first week of the class. Evaluation will be based on several individual assignments and exams.

Students must attain an overall passing grade on the weighted average of exams in the course in order to obtain a clear pass (C- or better).

**Required Books:**

Data Mining: Concepts and Techniques (3rd Ed), Jiawei Han, Micheline Kamber, and Jian Pei, Morgan Kaufmann Publishers, 2011, 9780123814791

**Academic Honesty Statement:**
Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct Policies (http://www.sfu.ca/policies/gazette/student.html).