

Computer Science (COMP) 682

Data Mining (Revision 3)

Status: Replaced with new revision, see the [course listing](#) for the current revision ✖

Delivery mode: Individualized study online

Credits: 3

Area of study: Information Systems

Prerequisites: COMP 602 or equivalent. Students registering in this course will need to have some background in database systems and statistics. Students who are concerned about not meeting the prerequisites for this course are encouraged to contact the course coordinator before registering.

Precluded: None

Faculty: Faculty of Science and Technology

Notes: This is a graduate level course and students need to apply and be approved to one of the graduate programs or as a non-program **School of Computing and Information Systems** graduate student in order to take this course. Minimum admission requirements must be met. Undergraduate students who do not meet

admission requirements will not normally be permitted to take this course.

Instructor:

Dr. Larbi Esmahi

Overview

Our ability to generate and collect data has been increasing rapidly. The widespread use of information technology in our lives has flooded us with a tremendous amount of data. This explosive growth of stored and transient data has generated an urgent need for new techniques and automated tools that can assist in transforming this data into useful information and knowledge. Data mining has emerged as a multidisciplinary field that addresses this need.

This course discusses techniques for preprocessing data before mining and presents the concepts related to data warehousing, online analytical processing (OLAP), and data generalization. It presents methods for mining frequent patterns, associations, and correlations. It also presents methods for data classification and prediction, data-clustering approaches, and outlier analysis.

Outline

Unit 1: Overview of Data Mining

- This unit provides some background on data objects and statistical concepts. It also discusses the types of data to be mined and presents a general classification of data-mining tasks.

Unit 2: Data Preprocessing

- This unit introduces techniques for preprocessing data before mining. Concepts such as the cleaning, integration, reduction, transformation, and discretization of data are discussed.

Unit 3: Overview of Data Warehousing and OLAP

- This unit provides a solid introduction to data warehousing, OLAP, and data generalization.

Unit 4: Data Cube Computation and Multidimensional Data Analysis

- This unit presents a detailed study of methods for data cube computation, advanced query processing, and multidimensional data analysis.

Unit 5: Mining Frequent Patterns, Associations, and Correlations

- This unit presents methods for mining frequent patterns, associations, and correlations.

Unit 6: Classification

- This unit discusses ways of classifying data: decision tree induction, Bayesian classification, rule-based classification, neural networks, support vector machines, associative classification, *k*-nearest-neighbor classifier, case-based reasoning, genetic algorithms, rough sets, and fuzzy set approaches.

Unit 7: Cluster Analysis

- This unit describes the partitioning, hierarchical, density-based, grid-based, and model-based methods data clustering.

Unit 8: Outlier Detection


- This unit describes several major approaches to the detection of anomalies, such as the statistical, proximity-based, clustering-based, and classification-based methods.

Learning outcomes

Upon successful completion of this course, you should be able to

- interpret the contribution of data warehousing and data mining to the decision-support level of organizations.
- evaluate different models used for OLAP and data preprocessing.
- categorize and carefully differentiate between situations for applying different data-mining techniques: frequent pattern mining, association, correlation, classification, prediction, cluster, and outlier analysis.
- design and implement systems for data mining.
- evaluate the performance of different data-mining algorithms.
- propose data-mining solutions for different applications.

Evaluation

To **receive credit** for COMP 682, you must achieve a cumulative course grade of **B- (70 percent)**  or better, and must achieve an average grade of at least 60% on the assignments and project and 60% on the final examination. Your cumulative course grade will be based on the following assessment.

Activity	Weight
Assignment 1	10%
Assignment 2	15%
Assignment 3	15%
Project	30%
Final Invigilated Examination	30%
Total	100%

The **final examination** for this course must be requested in advance and written under the supervision of an AU-approved exam invigilator. Invigilators include either ProctorU or an approved in-person invigilation centre that can accommodate online exams. Students are responsible for payment of any invigilation fees. Information on exam request deadlines, invigilators, and other exam-related questions, can be found at the **Exams and grades** section of the Calendar.

Materials

Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann, 2012. ISBN: 9780123814807. (eText)

Other References

Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann, 2011. ISBN 978-0-12-374856-0. (Available as an e-book through the Athabasca University Library.)

Other

The remainder of the course material is distributed through the online course site.

Important links

- [Future Course Offerings](#)
- [Important Dates and Deadlines](#)
- [MSc IS Contact Information](#)

Athabasca University reserves the right to amend course outlines occasionally and without notice. Courses offered by other delivery methods may vary from their individualized study counterparts.

Opened in Revision 3, August 11, 2017

Updated December 8, 2024

View [previous revision](#)
