ADA University, School of IT and Engineering

CSCI 4700 – Data Mining

Instructor: **Samir T. Mammadov**
Email: **stmammadov@ada.edu.az**
Office: **B315**
Office hours: **Tuesday/Thursday, 16:00-18:00**

Schedule: **Tuesday/Thursday, 11:30-12:45, A202**

Synopsis

Data mining enables one to gain fundamental insights and knowledge from data and allows the discovery of novel and insightful patterns in both structured and unstructured data. The course introduces the main topics in data mining and knowledge discovery, including algebraic and statistical foundations, pattern mining, rule learning, classification, and clustering. Emphasis is laid on the algorithmic approach with geometric and probabilistic viewpoints also being leveraged. Students will have an opportunity to work with various modern tools to finish term course assignments and projects. The course requires introductory knowledge of linear algebra, probability, and data structures.

Outcomes

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

- Understand what data mining is, what kinds of data can be mined, what kinds of patterns can be mined, and what kinds of applications are targeted.
- Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.
- Explain major issues in data mining.
- Apply machine learning and pattern recognition algorithms, as well as statistics in data mining applications.
- Identify what kinds of technologies are used for different applications.

Prerequisites

- CSCI 2304 – Data Structures & Algorithms
- MATH 2406 – Probability & Statistics
- MATH 3501 – Linear Algebra

Reading

Primary source(s):

- Zaki, M., Meira, W. *Data Mining and Machine Learning, 2nd edition.* 2020.

Supplementary source(s):

- Skiena, S. *The Data Science Design Manual.* 2017.
- Hastie, T., Tibshirani, R., Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition.* 2016.

Grading

- **Assignments** **30%**
- **Project** **20%**
- **Midterm** **20%**
- **Final** **30%**

Topics

The following list gives an overview of the core topics intended to be covered during the semester. This list is not absolute and due to changes, by adding or removing some topics, as the instructor considers necessary.

- **Foundations**
  - **Data Matrix**
  - **Numeric and Categorical Attributes**
  - **Graph Data**
  - **Data Visualization**
- **Frequent Pattern Mining**
  - **Itemset Mining**
  - **Summarizing Itemsets**
  - **Sequence Mining**
  - **Graph Pattern Mining**
  - **Pattern and Rule Assessment**
- **Clustering**
  - **Representative-based Clustering**
  - **Hierarchical Clustering**
  - **Clustering Validation**