

**Title (Units):** COMP4027 Data Mining and Knowledge Discovery (3,2,1)

**Course Aims:** To learn the latest development of knowledge discovery and data mining concepts and techniques. Theories and algorithms for data mining and knowledge discovery will be introduced. Relevant applications in specific domains such as medicine and health care will be covered.

**Prerequisite:** i) COMP 2015 Data Structures and Algorithms  
 COMP 2016 Database Management  
 MATH 2005 Calculus, Probability and Statistics for Computer Science  
 Or  
 ii) COMP 2865 Fundamental of Data Analysis and Management

Anti-requisite: MATH3836 Data Mining

**Course Intended Learning Outcomes (CILOs):**

Upon successful completion of this course, students should be able to:

| No. | Course Intended Learning Outcomes (CILOs)                       |
|-----|---|
|     | <b>Knowledge</b>  |
| 1   | Distinguish data mining applications from other IT applications |
| 2   | Explain data mining algorithms                                  |
| 3   | Explain applicability of data mining                            |
|     | <b>Professional Skill</b>                                       |
| 4   | Suggest appropriate solutions to data mining problems           |
| 5   | Analyze data mining algorithms and techniques                   |
|     | <b>Attitude</b>   |
| 6   | Work as a team in solving challenging data mining problems      |

**Calendar Description:** This course is aimed at providing an overview of concepts and techniques in knowledge discovery and data mining. Relevant applications in specific domains such as medicine and health care will be covered.

**Teaching and Learning Activities (TLAs):**

| CILOs | Type of TLA   |
|-------|---|
| 1-3   | Student will learn the concepts from lecture  |
| 4-5   | Student will achieve the outcomes via assignment on data mining                               |
| 4-5   | Student will achieve the outcomes via guided laboratory with data mining software             |
| 4-6   | Student will achieve the outcomes via group project on solving real world data mining problem |

**Assessment:**

| No. | Assessment Methods    | Weighting | CILOs to be addressed | Description of Assessment Tasks   |
|-----|-----------------------|-----------|-----------------------|---|
| 1   | Continuous Assessment | 40%       | 4-6                   | Assignments and labs will be used to consolidate their knowledge and develop their skills in data mining. Lab exercises and projects will further strengthen their understanding and problem solving skills, which takes 15% of an overall score. The written assignments and quizzes take 25% of an overall score. |
| 2   | Examination           | 60%       | 1-5                   | Final examination questions are designed to see how far students have achieved their intended learning outcomes. Analysis based questions will be used to assess the understanding of data mining problems. Problem solving questions will be used to   |

|  |  |  |  |   |
|--|--|--|--|---|
|  |  |  |  | assess students' ability in tackling applications in data mining. |
|--|--|--|--|---|

**Assessment Rubrics:**

|  | <b>Excellent (A)</b>  | <b>Good (B)</b>   | <b>Satisfactory (C)</b>                                | <b>Marginal Pass (D)</b>  | <b>Fail (F)</b>  |
|--|---|---|--|---|--|
| Identify and distinguish data mining applications from other IT applications | Thorough identification of almost all applications                    | Identification of a large number of applications              | Identification of a moderate number of applications    | Identification of a small number of applications                    | Identification of only a very small number of applications               |
| Describe data mining algorithms  | Thorough description of almost all data mining algorithms             | Description of most of the algorithms                         | Description of some of the algorithms                  | Description of a small number of algorithms                         | Description of only a few number of algorithms                           |
| Describe applicability of data mining  | Thorough description of almost all usage of data mining               | Description of most of the usage                              | Description of some of the usage                       | Description of a small number of usage                              | Description of a very small number of usage                              |
| Suggest appropriate solutions to data mining problems                        | Suggestion of almost all correct solutions                            | Suggestion of most of the solutions                           | Suggestion of some of the solutions                    | Suggestion of a small number of solutions                           | Suggestion of a very small number of solutions                           |
| Analyze data mining algorithms and techniques                                | Thorough analysis of almost all data mining algorithms and techniques | Analysis of most of the data mining algorithms and techniques | Analysis of some data mining algorithms and techniques | Analysis of a small number of data mining algorithms and techniques | Analysis of a very small number of data mining algorithms and techniques |

**Course Content and CIOs Mapping:**

| <b>Content</b> |                             | <b>CIO No.</b> |
|----------------|-----------------------------|----------------|
| I              | Introduction to Data Mining | 1              |
| II             | Data Mining Algorithms      | 2, 3, 5, 6     |
| III            | Clustering                  | 2, 5           |
| IV             | Case Studies in Data Mining | 3, 4, 5, 6     |

**References:**

- P. Tan, M. Steinback and V. Kumar, Introduction to Data Mining, Addison Wesley, Second Edition, 2016.
- J. Leskovec, A. Rajaraman, and J. Ullman, Mining of Massive Datasets, Second Edition, Cambridge University Press, 2014.
- I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, San Francisco, CA, Third Edition, 2011.
- J. W. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA, Third Edition, 2011.
- D. Hand, H. Mannila and P. Smyth, Principles of Data Mining, MIT Press, 2001

**Course Content:**

## **Topic**

- I. Introduction to Data Mining
  - A. Overview of data mining
  - B. Data preparation for knowledge discovery
  - C. Data warehousing
  - D. Data visualization and exploration
  
- II. Data Mining Algorithms
  - A. Association rules
  - B. Apriori algorithms
  - C. Classification algorithms
  - D. Mining event sequences
  - E. Applications
  
- III. Clustering
  - A. Classical algorithms
  - B. Graph-based clustering
  - C. Advanced methods
  - D. Applications
  
- IV. Case Studies in Data Mining
  - A. Health Informatics
  - B. Related applications