

**Jerash University**

**Faculty of Computer Science and Information Technology**

**Computer Sciences Department**

**Semester**: Fall Semester 2018/2019

|  |  |
| --- | --- |
| **Course symbol and number: 1002442** | **Course Name:** Data Mining |
| **Teaching Language:** English | **Prerequisites: 1002340**. |
| **Credits:** 3 hours**.** | **Course Level:** 400 |

|  |
| --- |
| **Course Description** |
| Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called *Knowledge Discovery* that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical, pattern recognition and machine learning techniques, and reporting and visualization of the generated structures. The course will cover all these issues and will illustrate the whole process by examples of practical applications.   The students will use recent Data Mining software. |

|  |
| --- |
| **Course Objectives** |
| * To provide both theoretical and practical coverage of all data mining topics. * To introduce students to the basic concepts and techniques of Data Mining. * To develop skills of using recent data mining software for solving practical problems.   Topics covered include; predictive modeling, association analysis, clustering, anomaly detection, visualizationwill implement solutions via C/C++ programs, and through NACHOS. |

|  |
| --- |
| **Learning Outcomes** |
| At the completion of the course, students will be able to…  1. **Apply** supervised learning algorithms to prediction problems and **evaluate** the results (ABET outcomes: A, B, C)  2. **Apply** unsupervised learning algorithms to data analysis problems and **evaluate** results (ABET outcomes: A, B, C)  3. **Apply** reinforcement learning algorithms to control problem and **evaluate** results (ABET outcomes: A, B, C)  4. **Decide** what kind of problem (supervised, unsupervised, or reinforcement) it is, given a description of a new problem (ABET outcomes: B, C) |

|  |  |
| --- | --- |
|  | **Text Book(s)** |
| **Title** | Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, |
| **Author(s)** | Witten, l. and Eibe, I., |
| **Publisher** | Morgan Kaufmann |
| **Year** | 2005 |
| **Edition** |  |

|  |  |
| --- | --- |
|  | **References** |
| **Books** | Witten, l. and Eibe, I., Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, 2005, Morgan Kaufmann |
| **Internet links** | http://www.jpu.edu.jo/lms |
| **Course link** |  |

|  |  |
| --- | --- |
|  | **Instructors** |
| **Instructor** | Dr.Ali Malkawi |
| **Office Location** | الطابق السابع - 715 |
| **Office Phone** |  |
| **E-mail** | [ali.amalkawi@jpu.edu.jo](mailto:Shquier@jpu.edu.jo) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Topics Covered** | | | |
| **Topics** | **Chapters in Text** | **Week number** | **Teaching hours** |
| **Introduction**  Data Flood  Data Mining Application Examples  Data Mining and Knowledge Discovery  Data Mining Tasks | **Chp.1** | *1* | *3* |
| **Machine Learning and Classification**  Machine Learning and Classification  Examples  Learning as Search  Bias  Weka | **Chp.1** | *2,3* | *12* |
| **Input: Concepts, instances, attributes**  What is a concept?  What is an example?  What is an attribute?  Preparing the data | **Chp. 2** | *4,5* | *12* |
|  |  | *6,7* | *12* |
| **Output: Knowledge Representation**  Decision tables  Decision trees  Decision rules  Rules involving relations  Instance-based representation | **Chp.3** |  |  |
| **Classification - Basic methods**  OneR  NaiveBayes | **Chp.4** | *8,9* | *12* |
| **Classification: Decision Trees**  Top-Down Decision Trees  Choosing the Splitting Attribute  Information Gain and Gain ratio | **Chp.4** | *10,11* | *12* |
|  |  | *12* | *6* |
| **Classification: C4.5**  Handling Numeric Attributes  Finding Best Split  Dealing with Missing Values  Pruning  Pre-pruning, Post-Pruning, Estimating Error Rates  From Trees to Rules | **Chp.5** | *13* | *6* |
| **Classification: more methods**  Rules  Regression  Instance-based (Nearest neighbor) | **Chp.4** | *14,15* | *12* |

|  |  |  |
| --- | --- | --- |
|  | ***Evaluation*** |  |
| **Assignment and Projects** |  | ***20%*** |
| Project |  | **15%** |
| Presentation & Discussion |  | **5%** |
| **Individual Work** |  | *80%* |
| Attendance, Participation, Home works and short report | Chapter Homework’s, Discussions, Short Presentations | **10%** |
| Quizzes | Unannounced Short quizzes |
| First Exam | Multiple Choice Questions worth 25% and Essay Questions worth 75% of exam grade. | **15%** |
| Second Exam | Multiple Choice Questions worth 25% of and Essay Questions worth 75% of exam grade. | **15%** |
| A Comprehensive Final examination | Multiple Choice Questions worth 25% and Essay Questions worth 75% of exam grade. | **40%** |
| *total* |  | *100%* |

|  |  |
| --- | --- |
|  | ***Policy*** |
| ***Attendance*** | *Attendance is very important for the course. In accordance with university policy, students missing more than the allowed absence rate of total classes are subject to failure. Penalties may be assessed without regard to the student's performance. Attendance will be recorded at the beginning or end of each class.* |
| ***Exams*** | *All exams will be CLOSE-BOOK; necessary algorithms/equations/relations will be supplied as convenient.* |

|  |
| --- |
| **Class Schedule & Room** |

|  |  |  |
| --- | --- | --- |
| **Office Hours** | | |
| Sun: 12 - 2  Mon: 8 - 11  Tues: 12 - 2  Wed: 8 - 11 | | |
|  | \* Or by an appointment through email |  |

|  |  |
| --- | --- |
|  | **Teaching Assistant** |
| To announced later on. |  |

|  |  |
| --- | --- |
|  | **Prerequisites** |
| **Prerequisites by course** | 1001108 |