MODULE DESCRIPTION FORM

نموذج وصف المادة الدراسية

| **Module Information**  **معلومات المادة الدراسية** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Module Title** | Data mining | | | | **Module Delivery** | | |
| **Module Type** | Core | | | | * **☒ Theory** * **☐ Lecture** * **☒ Lab** * **☐ Tutorial** * **☐ Practical** * **☐ Seminar** | | |
| **Module Code** | BMI412 | | | |
| **ECTS Credits** | 5.00 | | | |
| **SWL (hr/sem)** | 125 | | | |
| **Module Level** | | 4 | **Semester of Delivery** | | | | 7 |
| **Administering Department** | | BID | **College** | BMIC | | | |
| **Module Leader** | Mohammed Abdul Ameer Jabbar | | **e-mail** | [mohammedaji@uoitc.edu.iq](mailto:mohammedaji@uoitc.edu.iq) | | | |
| **Module Leader’s Acad. Title** | | Assis. Lecturer | **Module Leader’s Qualification** | | | | MSc |
| **Module Tutor** |  | | **e-mail** |  | | | |
| **Peer Reviewer Name** | | omar A.M | **e-mail** | omara.m@uoitc.edu.iq | | | |
| **Scientific Committee Approval Date** | | 18/6/2023 | **Version Number** | | | 1.0 | |

| **Relation with other Modules**  **العلاقة مع المواد الدراسية الأخرى** | | | |
| --- | --- | --- | --- |
| **Prerequisite module** | Pattern discovery in bioinformatics / BID321 | **Semester** | 6 |
| **Co-requisites module** | None | **Semester** |  |

| **Module Aims, Learning Outcomes and Indicative Contents**  **أهداف المادة الدراسية ونتائج التعلم والمحتويات الإرشادية** | |
| --- | --- |
| **Module Aims**  **أهداف المادة الدراسية** | 1. To understand data warehouse concepts, architecture, business analysis and tools. 2. To understand data pre-processing and data visualization techniques. 3. To study algorithms for finding hidden and interesting patterns in data. 4. To understand and apply various clustering techniques using tools. |
| **Module Learning Outcomes**  **مخرجات التعلم للمادة الدراسية** | Upon completion of the course, the students should be able to:   1. Design a Data warehouse system and perform business analysis with OLAP tools. 2. Apply suitable pre-processing and visualization techniques for data analysis. 3. Apply frequent pattern and association rule mining techniques for data analysis. 4. Apply appropriate classification and clustering techniques for data analysis. |
| **Indicative Contents**  **المحتويات الإرشادية** | **Indicative content includes the following:**  **UNIT I**  **DATA WAREHOUSING, BUSINESS ANALYSIS AND ON-LINE ANALYTICAL PROCESSING (OLAP)**  Basic Concepts - Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors - Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP.  **UNIT II**  **DATA MINING – INTRODUCTION**  Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.  **UNIT III**  **CLUSTER ANALYSIS and CLASSIFICATION**  Decision Tree Induction - Bayesian Classification – Rule Based Classification -Techniques to improve Classification Accuracy. Clustering Techniques – Cluster Analysis - Partitioning Methods - Hierarchical Methods – Density Based Methods - Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods. |

| **Learning and Teaching Strategies**  **استراتيجيات التعلم والتعليم** | |
| --- | --- |
| **Strategies** | The main strategies are covering All data mining issues including theoretically and practically in this course. It also includes several well-integrated examples and images. Every key topic is divided into three chapters. The first covers basic principles that serve as a foundation for understanding each data mining technique, and the second covers more advanced concepts and algorithms. The third includes implementation of data mining techniques into real medical projects. |

| **Student Workload (SWL)**  **الحمل الدراسي للطالب** | | | |
| --- | --- | --- | --- |
| **Structured SWL (h/sem)**  **الحمل الدراسي المنتظم للطالب خلال الفصل** | 64 | **Structured SWL (h/w)**  **الحمل الدراسي المنتظم للطالب أسبوعيا** | 4 |
| **Unstructured SWL (h/sem)**  **الحمل الدراسي غير المنتظم للطالب خلال الفصل** | 61 | **Unstructured SWL (h/w)**  **الحمل الدراسي غير المنتظم للطالب أسبوعيا** | 4 |
| **Total SWL (h/sem)**  **الحمل الدراسي الكلي للطالب خلال الفصل** | 125 | | |

| **Module Evaluation**  **تقييم المادة الدراسية** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **As** | | **Time/Number** | **Weight (Marks)** | **Week Due** | **Relevant Learning Outcome** |
| **Formative assessment** | **Quizzes** | 2 | 10% (10) | 6, 11 | LO #1, 2 |
| **Assignments** | 2 | 10% (10) | 4, 10 | LO # 3, 4, |
| **Projects / Lab.** | 1 | 10% (10) | Continuous |  |
| **Report** | 1 | 10% (10) | 13, 14 | LO # 2, 4 |
| **Summative assessment** | **Midterm Exam** | 2 hr | 10% (10) | 15 | LO # 1, 2, 3 |
| **Final Exam** | 3hr | 50% (50) | 16 | All |
| **Total assessment** | | | 100% (100 Marks) |  |  |

| **Delivery Plan (Weekly Syllabus)**  **المنهاج الاسبوعي النظري** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1** | Data Mining concepts and techniques - Introduction |
| **Week 2** | Knowledge Discovery in Databases (KDD) |
| **Week 3** | data warehouses  Data Warehouse Design Process  Data Warehouse operations |
| **Week 4** | Data Warehouse Models  Metadata Repository |
| **Week 5** | Online analytical Processing OLAP  Type of OLAP (MOLAP, ROLAP and HOLAP) the architectures and definitions |
| **Week 6** | Quiz 1 |
| **Week 7** | Cluster analyses  Clustering in data mining  Types of Clustering Methods |
| **Week 8** | Clustering Algorithms  Hierarchical Clustering and its types  Linkage Criteria or distance measures |
| **Week 9** | Partitioning and fuzzy method  K-Means algorithm  Fuzzy c-means |
| **Week 10** | Density Based method  DBSCAN Algorithm  Mean shift algorithm  Optical algorithm |
| **Week 11** | Quiz 2 |
| **Week 12** | Applications of data mining in medical aspects |
| **Week 13** | Projects discussion and Seminar |
| **Week 14** | Seminar reports |
| **Week 15** | **Mid term** |

| **Delivery Plan (Weekly Lab. Syllabus)**  **المنهاج الاسبوعي للمختبر** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1-2** | Lab1: how to install clustering method using python sikit- learn |
| **Week 3-4** | Lab2: select medical dataset with clustering issues to prepare the projects |
| **Week 5-6** | Lab3: data mining techniques implementation |
| **Week 7-8** | Lab4: K-Means, Fuzzy c-means, mean shift, optical, implementation in python code |
| **Week 9-10** | Lab5: DBSCAN, Affinity implementation in python code |
| **Week 11-12** | Lab6: model evaluation and similarity measures |
| **Week 13-15** | Lab7: practice |

| **Learning and Teaching Resources**  **مصادر التعلم والتدريس** | | |
| --- | --- | --- |
|  | **Text** | **Available in the Library?** |
| **Required Texts** |  | Yes |
| **Recommended Texts** | Jiawei Han and Micheline Kamber, ―Data Mining Concepts and Techniques‖, Third Edition, Elsevier, 2012. | No |
| **Websites** |  | |

| **Grading Scheme**  **مخطط الدرجات** | | | | |
| --- | --- | --- | --- | --- |
| **Group** | **Grade** | التقدير | **Marks (%)** | **Definition** |
| **Success Group**  **(50 - 100)** | **A -** Excellent | **امتياز** | 90 - 100 | Outstanding Performance |
| **B -** Very Good | **جيد جدا** | 80 - 89 | Above average with some errors |
| **C -** Good | **جيد** | 70 - 79 | Sound work with notable errors |
| **D -** Satisfactory | **متوسط** | 60 - 69 | Fair but with major shortcomings |
| **E -** Sufficient | **مقبول** | 50 - 59 | Work meets minimum criteria |
| **Fail Group**  **(0 – 49)** | **FX –** Fail | **راسب (قيد المعالجة)** | (45-49) | More work required but credit awarded |
| **F –** Fail | **راسب** | (0-44) | Considerable amount of work required |
|  |  |  |  |  |
| **Note:** Marks Decimal places above or below 0.5 will be rounded to the higher or lower full mark (for example a mark of 54.5 will be rounded to 55, whereas a mark of 54.4 will be rounded to 54. The University has a policy NOT to condone "near-pass fails" so the only adjustment to marks awarded by the original marker(s) will be the automatic rounding outlined above. | | | | |