|  |  |
| --- | --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **Trường Đại học GTVT**  Khoa: Công nghệ thông tin | **MINISTRY OF EDUCATION AND TRAINING**  **University of Transport and Communications**  Faculty of Information Technology |

ĐỀ CƯƠNG HỌC PHẦN

**KHAI PHÁ DỮ LIỆU**

**Data mining**

**1. Thông tin tổng quát** *(general information)*

|  |  |
| --- | --- |
| * Tên học phần: | Khai phá dữ liệu |
| * Mã số học phần: | IT1.231.3 |
| * Ngành/Chuyên ngành đào tạo | Công nghệ thông tin, Khoa học máy tính |
| * Thuộc khối kiến thức/ kỹ năng:   🞎 Kiến thức cơ bản 🞎 Kiến thức chuyên ngành  🞎 Kiến thức cơ sở ngành 🗹 Kiến thức ngành | |
| * Số tín chỉ: | 3 |
| + Số tiết lý thuyết: | 30 |
| + BTL |  |
| + Số tiết Thảo luận, Bài tập: | 15 |
| + Số tiết, thực hành, thí nghiệm: | 15 |
| + Số tiết tự học: | 90 |
| * Học phần tiên quyết: |  |
| * Học phần học trước: |  |
| * Học phần song hành: |  |
| * Yêu cầu khác đối với học phần: | Phòng học có máy chiếu/Bảng thông minh |

*(Số tiết phân bổ cho lý thuyết, thảo luận, bài tập, bài tập lớn, thực hành, thí nghiệm điền và bôi đậm tùy theo học phần cụ thể)*

**2. Mô tả học phần***(course descriptions)*

- Học phần thuộc khối kiến thức cơ sở ngành.

- Học phần cung cấp cho sinh viên kiến thức cơ bản về khai phá dữ liệu : giới thiệu quá trình khám phá tri thức, các khái niệm, công nghệ, và ứng dụng của khai phá dữ liệu. Ngoài ra, học phần này cũng trình bày các vấn đề tiền xử lý dữ liệu, các bước trong khai phá dữ liệu, các giải thuật và công cụ khai phá dữ liệu mà có thể được dùng hỗ trợ nhà phân tích dữ liệu và nhà phát triển ứng dụng khai phá dữ liệu.

- Sinh viên có khả năng làm việc nhóm để khảo sát, hình thành ý tưởng và xây dựng ứng dụng khai phá dữ liệu trong bài toán thực tế.

**3. Nguồn học liệu** *(learning resources: course books, reference books, and softwares)*

*(Các giáo trình, tài liệu tham khảo, các phần mềm, không quá 5 cuốn)*

***Giáo trình:***

***Tài liệu khác:***

[1]. Hà Quang Thụy, Giáo trình Khai phá dữ liệu Web, NXB Giáo dục Việt Nam, 2011.

[2]. Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, Third Edition, Morgan Kaufmann Publishers, 2011.

**4. Mục tiêu học phần** *(course goals)*

*(Thể hiện kiến thức, kỹ năng mà môn học cung cấp và sự liên quan với CĐR của CTĐT được phân nhiệm cho môn học)*

|  |  |  |
| --- | --- | --- |
| **Mục tiêu (G.x) [1]** | **Mô tả mục tiêu [2]** | **CĐR liên quan của CTĐT [3]** |
| **G.1** | **Nắm được** kiến thức tổng quan về khai phá dữ liệu và **sử dụng** được các kỹ thuật khai phá dữ liệu trong bài toán cụ thể. | CĐR2 (1.3) |
| **G.2** | **Áp dụng** các kỹ thuật khai phá dữ liệu để xây dựng ứng dụng khai phá dữ liệu cho một bài toán thực tế. | CĐR4 (4.6)  CĐR9 (4.1) |

*[1]: Ký hiệu mục tiêu của học phần. [2]: Mô tả mục tiêu. [3]: Ký hiệu CĐR của CTĐT và chuẩn đầu ra CDIO tương ứng.*

**5. Chuẩn đầu ra học phần** *(course learning outcomes)*

*(CĐR chi tiết hơn mục tiêu, mô tả sau khi học xong* ***sinh viên sẽ đạt được gì*** *về kiến thức, kỹ năng và thái độ)*

|  |  |  |
| --- | --- | --- |
| **CĐR HP cấp độ 3**  **(G.x.y) [1]** | **Mô tả CĐR học phần[2]** | **Mức độ chung HP theo Bloom [3]** |
| **G.1.1** | **Hiểu** được các khái niệm cơ bản và giải thích được các thuật ngữ trong khai phá dữ liệu(BL3) | 1.3 – TUA3 |
| **G.1.2** | **Nắm bắt** đượcquy trình và kỹ thuật khai phá dữ liệu(BL3) |
| **G.1.3** | **Hiểu** được các thuật toán khai phá dữ liệu (BL3) |
| **G.2.1** | **Sử dụng** kỹ năng phân tích lựa chọn được kỹ thuật khai phá dữ liệu thích hợp với yêu cầu của bài toán (BL3) | 4.1 (5-6) – UA4 |
| **G.2.2** | **Áp dụng** kỹ thuật khai phá dữ liệu cho bài toán thực tế (BL3) | 4.6.4 – UA4 |

*[1]: Ký hiệu CĐR của học phần. [2]: Mô tả CĐR học phần, bao gồm các động từ chủ động theo Bloom’s Taxonomy, khuyến khích viết tích hợp kỹ năng và kiến thức [3]: Mức độ năng lực mà HP đảm trách theo hoạt động TUA.*

**6. Đánh giá học phần** *(course assessment)*

*(các thành phần, các bài đánh giá, các tiêu chí đánh giá, chuẩn đánh giá, và tỷ lệ đánh giá, thể hiện sự liên quan với các CĐR của học phần)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Thành phần đánh giá [1]** | **Hình thức đánh giá (A.x.y) [2]** | **CĐR học phần (G.x.y) [3]** | **Tiêu chí**  **đánh giá [4]** | **Tỷ lệ (%)[5]** |
| A1. Đánh giá quá trình | A.1.1 Bài tập thực hành, điểm danh | G.1.1, 1.2, 1.3  G.2.1 | Chuyên cần | 10 |
| A.1.2 Bài kiểm tra giữa kỳ | G.1.2, 1.2  G2.1 | **Vận dụng** được các phương pháp tiền xử lý dữ liệu, các thuật toán khai phá dữ liệu (BL3) | 30 |
| A2. Đánh giá kết thúc học phần ít nhất 50%) | A.2.1 Bài thi cuối kỳ dưới dạng thi viết | G.1.2, G.1.3  G.2.2 | **Sử dụng** thành thạo các kỹ thuật tiền xử lý dữ liệu (BL3)  **Vận dụng** được các kỹ thuật khai phá dữ liệu (BL3) | 60 |

*[1]: Các thành phần đánh giá của học phần. [2]: Các bài đánh giá. [3]: Các CĐR được đánh giá. [4]: Tiêu chí đánh giá. [5]: Tỷ lệ điểm của các bài đánh giá trong tổng điểm học phần.*

**7. Khung kế hoạch giảng dạy***:*

*(Các nội dung giảng dạy theo chương mục, thể hiện sự liên quan với các CĐR của học phần và các bài đánh giá của học phần. Việc giảng dạy kỹ năng trong môn học có thể dạy lý thuyết và áp dụng hoặc có thể học qua trải nghiệm)*.

| **TT** | **Nội dung [2]** | **Số tiết** | **CĐR HP [3]** | **Hoạt động dạy và học [4]** | **Đánh giá [5]** |
| --- | --- | --- | --- | --- | --- |
| **1** | **Chương 1. Tổng quan về khai phá dữ liệu**  1.1 Giới thiệu chung  1.2 Khai phá dữ liệu là gì ?  1.3 Quá trình phát hiện tri thức trong cơ sở dữ liệu  1.4 Các kỹ thuật áp dụng trong khai phá dữ liệu  1.5 Ứng dụng của khai phá dữ liệu | 3 LT | G.1.1 | **Giảng viên:**  Kết hợp áp dụng một số hoạt động giảng dạy:  - Thuyết giảng (Lecturing) kết hợp với hỏi khái quát (Concept questions): Giảng dạy các nội dung lý thuyết, đặt các câu hỏi để kiểm tra, đánh giá sự tiếp thu của sinh viên  **Sinh viên:**  - Trong lớp: Nghe giảng lý thuyết, tham gia thảo luận theo sự phân nhiệm, tổ chức của giảng viên | A.1.1 |
| **2** | **Chương 2. Tiền xử lý dữ liệu**  2.1 Tầm quan trọng của tiền xử lý dữ liệu  2.2 Kỹ thuật làm sạch dữ liệu  2.3 Kỹ thuật tích hợp và chuyển đổi dữ liệu  2.4. Kỹ thuật giảm kích thước dữ liệu | 6 LT  3 BT  3 TH | G.1.2  G.1.3 | **Giảng viên:**  - Thuyết giảng (Lecturing) giảng dạy 2.1, 2.2,2.3,2.4  - Kết hợp Peer-to-Peer teaching, Team-based Learning và Online, Mobile and Blenđe Learning: Giao bài tập, hướng dẫn sinh viên làm bài tập, cung cấp link tài liệu online  - Tổ chức phân nhóm và giao bài tập lớn cho sinh viên.  **Sinh viên:**  - Trong lớp: Nghe giảng lý thuyết, tham gia thảo luận theo sự tổ chức của giảng viên.  - Ngoài lớp: Đọc trước slide, tài liệu giảng viên cung cấp theo sự phân công của giảng viên, trao đổi thảo luận với bạn cùng nhóm, làm bài tập theo sự hướng dẫn của giảng viên. | A.1.1  A.1.2  A.1.3  A.2.1 |
| **3** | **Chương 3. Khai phá luật kết hợp**  3.1 Khái niệm cơ bản  3.2 Thuật toán Apriori  3.3 Thuật toán FP-Growth | 6 LT  3 BT  3 TH | G.1.2  G.1.3 G.2.2 | **Giảng viên:**  - Thuyết giảng (Lecturing) giảng dạy các nội dung 3.1, 3.2, 3.3; sử dụng tài nguyên online nêu ví dụ áp dụng  - Kết hợp Peer-to-Peer teaching, Team-based Learning và Online, Mobile and Blenđe Learning: Giao bài tập, hướng dẫn sinh viên làm bài tập, cung cấp link tài liệu online  **Sinh viên:**  - Trong lớp: Nghe giảng lý thuyết, viết chương trình theo demo, tham gia thảo luận theo sự tổ chức của giảng viên, trình bày về các nội dung được phân công tìm hiểu, trình bày kết quả việc làm bài tập theo yêu cầu của giảng viên, thực hiện phát triển yêu câu và thực hành theo yêu cầu của giảng viên.  - Ngoài lớp: Đọc trước slide, tài liệu giảng viên cung cấp theo sự phân công của giảng viên, trao đổi thảo luận với bạn cùng nhóm, làm bài tập théo sự hướng dẫn của giảng viên, thực hành theo demo, phát triển yêu cầu và vận dụng kiến thức để giải quyết yêu cầu. | A.1.1  A.1.2  A.1.3  A.2.1 |
| **4** | **Chương 4: Phân lớp dữ liệu**  4.1 Các khái niệm cơ bản  4.2 Các loại độ đo và phương pháp đánh giá  4.3 Một số thuật toán phân lớp (Naïve Bayes, Cây quyết định) | 9 LT  6 BT  3 TH | G.1.2  G.1.3 G.2.2 | **Giảng viên:**  - Thuyết giảng (Lecturing) giảng dạy các nội dung 4.1, 4.2, 4.3; demo trên máy tính; sử dụng tài nguyên online nêu ví dụ áp dụng  - Kết hợp Peer-to-Peer teaching, Team-based Learning và Online, Mobile and Blenđe Learning: Giao bài tập, hướng dẫn sinh viên làm bài tập, cung cấp link tài liệu online  **Sinh viên:**  - Trong lớp: Nghe giảng lý thuyết, tham gia thảo luận theo sự tổ chức của giảng viên, trình bày về các nội dung được phân công tìm hiểu, trình bày kết quả việc làm bài tập theo yêu cầu của giảng viên.  - Ngoài lớp: Đọc trước slide, tài liệu giảng viên cung cấp theo sự phân công của giảng viên, trao đổi thảo luận với bạn cùng nhóm, làm bài tập theo sự hướng dẫn của giảng viên. | A.1.1  A.1.2  A.1.3  A.2.1 |
| **5** | **Chương 5: Phân cụm dữ liệu**  5.1 Các khái niệm cơ bản  5.2 Các loại độ đo cơ bản dùng trong phân cụm  5.3 Một số thuật toán phân cụm (k-means, k-medoid,) | 6 LT  3 BT  6 TH | G.1.2  G.1.3 G.2.2 | **Giảng viên:**  - Thuyết giảng (Lecturing) các nội dung 5.1, 5.2,5.3; sử dụng tài nguyên online nêu ví dụ áp dụng  - Kết hợp Peer-to-Peer teaching, Team-based Learning và Online, Mobile and Blenđe Learning: Giao bài tập, hướng dẫn sinh viên làm bài tập, cung cấp link tài liệu online  **Sinh viên:**  - Trong lớp: Nghe giảng lý thuyết, trình bày kết quả việc làm bài tập theo yêu cầu của giảng viên.  - Ngoài lớp: Đọc trước slide, tài liệu giảng viên cung cấp theo sự phân công của giảng viên, làm bài tập théo sự hướng dẫn của giảng viên. | A.1.1  A.1.2  A.1.3  A.2.1 |

*[1]: Thông tin về tuần/ buổi học. [2]: Liệt kê nội dung giảng dạy theo chương, mục. [3]: Liệt kê CĐR liên quan của học phần (ghi ký hiệu Gx.y). [4]: Mô tả chung các hoạt động dạy và học. [5]: Liệt kê các bài đánh giá liên quan (ghi ký hiệu Ax.y).*

**8. Quy định của học phần** *(course requirements and expectations)*

*(các quy định của học phần (nếu có), thí dụ: sinh viên không nộp bài tập và báo cáo đúng thời hạn, được coi như không nộp bài; sinh viên vắng 2 buổi thực hành trở lên, không được phép dự thi cuối kỳ…)*

**9. Phụ trách học phần**

- Bộ môn: Mạng và Các Hệ thống thông tin

- Địa chỉ và email liên hệ: P.308-A9, email: bmmht@utc.edu.vn

|  |  |
| --- | --- |
| **TRƯỞNG KHOA**  **TS. Hoàng Văn Thông** | **TRƯỞNG BỘ MÔN**  **TS. Nguyễn Quốc Tuấn** |

**PHÊ DUYỆT CỦA NHÀ TRƯỜNG**