|  |  |  |
| --- | --- | --- |
| http://www.tlu.edu.vn/Portals/0/2014/Logo-WRU.png | **TRƯỜNG ĐẠI HỌC THỦY LỢI**  KHOA CÔNG NGHỆ THÔNG TIN  BỘ MÔN HỆ THỐNG THÔNG TIN | **ĐỀ CƯƠNG HỌC PHẦN**  Trình độ đào tạo: Đại học |

**KHAI PHÁ DỮ LIỆU**

Data mining

**Mã số**: CSE404

1. **Số tín chỉ**: 3 tín chỉ (LT: 2; TH/BT/TL: 1)
2. **Số tiết**: Tổng: 45;

Trong đó: LT: 30 ; BT/TH: 15 ; TN .... ; ĐA: .... ; BTL: ; TQ, TT: .... ;

1. **Thuộc chương trình đào tạo ngành:**

*- Học phần bắt buộc cho ngành:* Hệ thống thông tin

*- Học phần tự chọn cho ngành:* Công nghệ thông tin, Kỹ thuật phần mềm.

1. **Phương pháp đánh giá**:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hình thức** | **Số lần** | **Mô tả** | **Thời gian** | **Trọng số** |
| **1. ĐIỂM QUÁ TRÌNH (QT)** | | | | **0,4** |
| Chuyên cần, xây dựng bài | 1 | Đi học đầy đủ,  Thái độ học tập tích cực | - Suốt quá trình học | 25% |
| Kiểm tra 1 | 1 | Bài kiểm tra ngắn về tiền xử lý dữ liệu | - Tuần 7 | 25% |
| Kiểm tra 2 | 1 | Kiểm tra giữa kỳ về hồi quy, phân lớp, phân cụm dữ liệu, luật kết hợp. | - Tuần 12 | 50% |
| **2. ĐIỂM THI KẾT THÚC HỌC PHẦN (THM)** | | | | **0,6** |
| Thi vấn đáp + Bài tập lớn | 1 | Sinh viên trình bày kết quả Bài tập theo nhóm đã đăng ký. | 1-2 tuần sau khi kết thúc môn học |  |

**Điểm môn học = QT x 0.4 + THM x 0.6**

**5. Điều kiện ràng buộc học phần:**

*- Học phần tiên quyết* : ...................................................................................................

*- Học phần học trước* : Cơ sở dữ liệu, Hệ quản trị CSDL

*- Học phần song hành:*

*- Ghi chú khác****:***

**6. Nội dung tóm tắt học phần:**

***Tiếng Việt*** :

Môn học cung cấp kiến thức tổng quan về khai phá dữ liệu, giới thiệu các phương pháp tiền xử lý dữ liệu, các phương pháp thường dùng trong khai phá dữ liệu, bao gồm: hồi quy dữ liệu, phân lớp và dự đoán dữ liệu, phân cụm dữ liệu và khai phá luật kết hợp. Ngoài ra, môn học cũng giới thiệu về các ứng dụng và các công cụ dùng trong khai phá dữ liệu. Sinh viên hiểu và sử dụng được các công cụ khai phá dữ liệu trên các bài toán ứng dụng trong thực tế, có thể phát triển các ứng dụng nâng cao xung quanh các tác vụ khai phá dữ liệu.

***Tiếng Anh*** :

The course provides an overview of data mining, introduction of data preprocessing methods, methods commonly used in data mining, including: data regression, classification and data prediction data, data clustering and association rules. In addition, the course also introduces applications and tools used in data mining. Students understand and use data mining tools on practical application problems, can develop advanced applications around data mining tasks.

**7. Cán bộ tham gia giảng dạy:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TT** | **Họ và tên** | **Học hàm, học vị** | **Điện thoại liên hệ** | **Email** | **Chức danh, chức vụ** |
| 1 | Đặng Thị Thu Hiền | TS |  | hiendt@tlu.edu.vn | GVC, Trưởng BM |
| 2 | Nguyễn Huy Đức | TS | 0903402655 | ducnghuy@tlu.edu.vn | GVCC |
| 3 | Trần Mạnh Tuấn | TS |  | tmtuan@tlu.edu.vn | Giảng viên, Phó BM |
| 4 | Nguyễn Tu Trung | TS |  | trungnt@tlu.edu.vn | Giảng viên |
| 5 | Nguyễn Ngọc Quỳnh Châu | ThS |  | chaunnq@tlu.edu.vn | Giảng viên |
| 6 | Lương Thị Hồng Lan | TS |  | lanlhbk@tlu.edu.vn | Giảng viên |

**8. Giáo trình sử dụng, tài liệu tham khảo:**

***Giáo trình:***

[1]. Slides bài giảng của Giảng viên.

[2]. Jiawei Han, Micheline Kamber, Jian Pei, *“Data Mining: Concepts and Techniques”*, Third Edition, Morgan Kaufmann Publishers, 2012.

***Các tài liệu tham khảo:***

[1]. David Hand, Heikki Mannila, Padhraic Smyth, “*Principles of Data Mining*”, MIT Press, 2001.

[2]. David L. Olson, Dursun Delen, “*Advanced Data Mining Techniques*”, Springer-Verlag, 2008.

[3]. Graham J. Williams, Simeon J. Simoff, “*Data Mining: Theory, Methodology, Techniques, and Applications*”, Springer-Verlag, 2006.

[4]. Hillol Kargupta, Jiawei Han, Philip S. Yu, Rajeev Motwani, and Vipin Kumar, “*Next Generation of Data Mining*”, Taylor & Francis Group, LLC, 2009.

[5]. Daniel T. Larose, “*Data mining methods and models*”, John Wiley & Sons, Inc, 2006.

[6]. Ian H.Witten, Frank Eibe, Mark A. Hall, “*Data mining: practical machine learning tools and techniques*”, Third Edition, Elsevier Inc, 2011.

**9. Nội dung chi tiết:**

**9.1 Lý thuyết (30 tiết)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TT** | **Nội dung (1)** | **Hoạt động dạy và học (2)** | **Số tiết** | | |
| LT | BT | TH/  TN/  TQ |
| 1 | Giới thiệu học phần | \* Giảng viên:   * Giới thiệu các thông tin cá nhân để sinh viên nắm được và có thể liên lạc. * Giới thiệu môn học, nội dung môn học, hình thức kiểm tra, đánh giá kết quả và thi hết học phần. * Hướng dẫn, truyền đạt cho sinh viên kinh nghiệm và phương pháp học tập để đạt kết quả tốt. | 1 |  |  |
| 2 | **Chương 1 – Tổng quan về tiêng khai phá dữ liệu**  1.1. Khái niệm 1.1. Quá trình khám phá tri thức 1.2. Các khái niệm 1.3. Ý nghĩa và vai trò của khai phá dữ liệu 1.4. Ứng dụng của khai phá dữ liệu 1.5. Những vấn đề chính trong khai phá dữ liệu 1.6. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Đặt câu hỏi thắc mắc về môn học nếu cần thiết | 2 |  |  |
| 3 | **Chương 2 - Tiền xử lý dữ liệu**  2.1. Tổng quan về giai đoạn tiền xử lý dữ liệu 2.2. Tóm tắt mô tả về dữ liệu 2.3. Làm sạch dữ liệu 2.4. Tích hợp dữ liệu 2.5. Biến đổi dữ liệu 2.6. Thu giảm dữ liệu 2.7. Rời rạc hóa dữ liệu 2.8. Tạo cây phân cấp ý niệm 2.9. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa. * Ra bài tập về tiền xử lý dữ liệu   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Làm bài tập theo yêu cầu của GV. | 3 |  | 1 |
| 4 | **Chương 3 – Hồi quy dữ liệu**  3.1. Tổng quan về hồi qui 3.2. Hồi qui tuyến tính 3.3. Hồi qui phi tuyến 3.4. Ứng dụng 3.5. Các vấn đề với hồi qui 3.6. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa. * Ra bài tập về hồi qui tuyến tính.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Thảo luận, làm bài tập theo yêu cầu. | 4 |  | 2 |
| 5 | **Chương 4 – Phân lớp dữ liệu**  4.1. Tổng quan về phân lớp dữ liệu 4.2. Phân lớp dữ liệu với cây quyết định 4.3. Phân lớp dữ liệu với mạng Bayesian 4.4. Phân lớp dữ liệu với mạng Neural 4.5. Các phương pháp phân lớp dữ liệu khác 4.6. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa. * Ra bài tập về phân lớp dữ liệu.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Thảo luận, làm bài tập theo yêu cầu. | 4 |  | 3 |
| 6 | **Chương 5 – Phân cụm dữ liệu**  5.1. Tổng quan về phân cụm dữ liệu 5.2. Phân cụm dữ liệu bằng phân hoạch 5.3. Phân cụm dữ liệu bằng phân cấp 5.4. Phân cụm dữ liệu dựa trên mật độ 5.5. Phân cụm dữ liệu dựa trên mô hình 5.6. Các phương pháp phân cụm dữ liệu khác 5.7. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa. * Ra bài tập về phân cụm dữ liệu.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Thảo luận, làm bài tập theo yêu cầu. | 5 |  | 3 |
| 7 | **Chương 6 – Luật kết hợp**  6.1. Tổng quan về khai phá luật kết hợp 6.2. Biểu diễn luật kết hợp 6.3. Khám phá các mẫu thường xuyên 6.4. Khám phá các luật kết hợp từ các mẫu thường xuyên 6.5. Khám phá các luật kết hợp dựa trên ràng buộc 6.6. Phân tích tương quan 6.7. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu hình ảnh/video minh họa. * Ra bài tập về tìm tập phổ biến, tìm luật kết hợp.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Giải quyết tình huống * Thảo luận, làm bài tập theo yêu cầu. | 5 |  | 2 |
| 8 | **Chương 7 – Công nghệ khai phá dữ liệu**  7.1. Tổng quan về công nghệ cơ sở dữ liệu  7.2. Khả năng hỗ trợ khai phá dữ liệu của công nghệ cơ sở dữ liệu.  7.3. Các ngôn ngữ truy vấn dành cho khai phá dữ liệu.  7.4. Hỗ trợ của các DBMS ngày nay dành cho khai phá dữ liệu.  7.5. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu minh họa.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Thảo luận, giải quyết tình huống. | 4 |  | 3 |
| 9 | **Chương 8 - Ứng dụng của khai phá dữ liệu**  8.1. Tổng quan về vấn đề phát triển ứng dụng khai phá dữ liệu 8.2. Qui trình phát triển ứng dụng khai phá dữ liệu 8.3. Các chuẩn dành cho khai phá dữ liệu 8.4. Các công cụ hỗ trợ phát triển ứng dụng khai phá dữ liệu 8.5. Ứng dụng minh họa 8.6. Tóm tắt | \* Giảng viên:   * Thuyết giảng * Truy vấn * Sử dụng trình chiếu minh họa.   \* Sinh viên:   * Trả lời các câu hỏi truy vấn * Thảo luận, giải quyết tình huống. | 2 |  | 1 |

**9.2 Thực hành (15 tiết)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Buổi** | **Số tiết** | **Nội dung** | **GHI CHÚ** |
| 1 | 3 | - Làm quen với một số môi trường/phần mềm thực hành: R, weka,....  - Làm một số bài về tiền xử lý dữ liệu |  |
| 2 | 3 | - Tiền xử lý DL (tiếp)  - Hồi quy dữ liệu |  |
| 3 | 3 | - Hồi quy DL  - Phân lớp DL |  |
| 4 | 3 | - Phân lớp DL  - Phân cụm DL  **- Kiểm tra giữa kỳ: 20-30 phút trên máy** |  |
| 5 | 3 | - Phân cụm DL  - Luật kết hợp |  |

**10. Chuẩn đầu ra (CĐR) của học phần:**

|  |  |  |
| --- | --- | --- |
| **STT** | **CĐR của học phần** | **CĐR của CTĐT**  **tương ứng** |
| 1 | **Kiến thức:**  - Nắm được nội dung các bước cơ bản của quá trình phát hiện tri thức từ dữ liệu. Phân biệt được phương pháp khai phá dữ liệu (KPDL) với các phương pháp xử lý dữ liệu truyền thống. Nắm được xu thế phát triển hiện đại của KPDL.  - Hiểu và vận dụng được: Các phương pháp tiền xử lý dữ liệu; Hồi quy dữ liệu; Các thuật toán điển hình trong phân lớp, phân cụm dữ liệu, tìm tập phổ biến và luật kết hợp; Các lỗi điển hình khi thực hiện dự án KPDL và phương hướng khắc phục; Sử dụng được một số nguồn tài nguyên thông dụng về KPDL | HTTT: 4  CNTT: 3,4,9  KTPM: 3 |
| 2 | **Kỹ năng:**  - Sử dụng và ứng dụng được các phương pháp, kỹ thuật và thuật toán khai phá dữ liệu phổ biến để giải quyết các vấn đề thực tế.  - Tìm được các giải pháp thích hợp và đúng đắn cho các vấn đề khai phá dữ liệu trong thế giới thực. | HTTT: 8,10,12  CNTT: 11  KTPN: 9 |
| 3 | **Năng lực tự chủ và trách nhiệm:**  Sẵn sàng tiếp cận và sử dụng công nghệ mới trong phát triển ứng dụng khai phá dữ liệu. | HTTT: 15 |
| 4 | **Phẩm chất đạo đức cá nhân, nghề nghiệp, xã hội**:  Có đạo đức, lương tâm nghề nghiệp, ý thức tổ chức kỷ luật và trách nhiệm với công việc, cộng đồng và xã hội. | CNTT: 22, 23 |

**11. Thông tin liên hệ của Bộ môn**

A. Địa chỉ bộ môn: Phòng 201 – Nhà c1, Trường Đại học Thủy lợi

B. Trưởng bộ môn: *(có trách nhiệm trả lời thắc mắc của sinh viên và các bên liên quan)*

- Họ và tên: TS. Đặng Thị Thu Hiền

- Số điện thoại: 0888288807

- Email: [hiendt@tlu.edu.vn](mailto:hiendt@tlu.edu.vn)

*Hà Nội, ngày tháng năm 2019*

|  |  |  |
| --- | --- | --- |
| **TRƯỞNG KHOA**  *(Phụ trách ngành đào tạo)*  **TS. Nguyễn Thanh Tùng** | **TRƯỞNG KHOA**  *(Phụ trách học phần)*  **TS. Nguyễn Thanh Tùng** | **TRƯỞNG BỘ MÔN**  **TS. Đặng Thị Thu Hiền** |