

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN: KHAI PHÁ DỮ LIỆU

Ngành đào tạo: Công nghệ thông tin

1. Thông tin tổng quát:

1.1. Thông tin về giảng viên

Giảng viên 1:

Họ và tên: Phan Anh Phong
Chức danh, học hàm, học vị: GVC. Tiến sĩ
Thời gian, địa điểm làm việc: Viện Kỹ thuật và Công nghệ, Đại học Vinh
Địa chỉ liên hệ: 182 Lê Duẩn, Thành phố Vinh, Tỉnh Nghệ An
Điện thoại, email: phongpa@vinhuni.edu.vn
Các hướng nghiên cứu chính: Các hệ cơ sở dữ liệu; Logic tính toán; Khai phá dữ liệu

Giảng viên 2:

Họ và tên: Cao Thanh Sơn
Chức danh, học hàm, học vị: GVC. Tiến sĩ
Thời gian, địa điểm làm việc: Viện Kỹ thuật và Công nghệ, Đại học Vinh
Địa chỉ liên hệ: 182 Lê Duẩn, Thành phố Vinh, Tỉnh Nghệ An
Điện thoại, email: sonct@vinhuni.edu.vn
Các hướng nghiên cứu chính: Web ngữ nghĩa; CSDL suy diễn; Khai phá dữ liệu

1.2. Thông tin về học phần:

| | |
|--|---------------------------------|
| - Tên học phần (tiếng Việt): Khai phá dữ liệu (tiếng Anh): Data mining | |
| - Mã số học phần: INF30045 | |
| - Thuộc khối kiến thức/kỹ năng: <input type="checkbox"/> Kiến thức cơ bản <input checked="" type="checkbox"/> Kiến thức chuyên ngành <input type="checkbox"/> Môn học chuyên về kỹ năng chung <input type="checkbox"/> Kiến thức cơ sở ngành <input type="checkbox"/> Kiến thức khác <input type="checkbox"/> Môn học đồ án tốt nghiệp | |
| - Số tín chỉ: | 03 |
| + Số tiết lý thuyết: | 30 |
| + Số tiết thảo luận/bài tập: | 0 |
| + Số tiết thực hành: | 15 |
| + Số tiết hoạt động nhóm: | 0 |
| + Số tiết tự học: | 90 |
| - Môn học tiên quyết: | - Không |
| - Môn học học trước: | - Công nghệ phần mềm (INF30034) |

2. Mô tả học phần

Học phần Khai phá dữ liệu (KPDL) cung cấp cho người học các khái niệm, nguyên lý, phương pháp, kỹ thuật và ứng dụng của khai phá dữ liệu, tập trung vào các thuật toán cơ bản của phân lớp, phân cụm và khai phá luật kết hợp như Apriori, ID3, Naïve Bayes, KNN,

K-means và phân cụm phân cấp... hướng tới cài đặt các hệ thống thông minh. Ngoài ra, học phân trang bị cho người học phương pháp tiếp cận, kỹ thuật đánh giá hiệu năng của các phương pháp khai phá dữ liệu.

3. Mục tiêu học phần

Sau khi học xong học phần này sinh viên có khả năng:

| Mục tiêu (Gx) (1) | Mô tả mục tiêu (2) | CDR của CTĐT (X.x.x) (3) | TĐNL (4) |
|-------------------|--|--------------------------|----------|
| G1 | <i>Phân tích</i> được quy trình và ứng dụng của khai phá dữ liệu | 1.4.1 | 4.0 |
| G2 | <i>Cài đặt</i> được các kỹ thuật khai phá dữ liệu cho tập dữ liệu mẫu | 1.4.1. | 4.0 |
| G3 | <i>Xây dựng</i> được các tiêu chí đánh giá và các kịch bản để triển khai khảo sát thử nghiệm | 2.2.3 | 3.5 |
| G4 | <i>Vận dụng</i> được các kiến thức liên quan trong thiết kế hệ thống thông minh | 4.3.3 | 4.0 |

4. Chuẩn đầu ra học phần

| Các chuẩn đầu ra học phần | | Trình độ năng lực | CDR CTĐT tương ứng | Mức độ giảng dạy |
|---------------------------|--|-------------------|--------------------|------------------|
| Ký hiệu | Nội dung CDR học phần | | | |
| G1.1 | <i>Chỉ ra</i> được các ứng dụng của khai phá dữ liệu trong các hoạt động kinh tế, y tế, giáo dục... | 4.0 | 1.4.1 | T |
| G1.2 | <i>Phân tích</i> được quy trình khi thực hiện khai phá dữ liệu | 4.0 | 1.4.1. | T |
| G1.3 | <i>Lựa chọn</i> được kỹ thuật tiền xử lý dữ liệu | 4.0 | 1.4.1 | TU |
| G2.1 | <i>Phân tích</i> được các bài toán về phân lớp, phân cụm và khai phá luật kết hợp | 4.0 | 1.4.1. | T |
| G2.2 | <i>Áp dụng</i> được các công cụ để thực hiện khai phá luật kết hợp, phân tích cụm và dự báo trên các tập dữ liệu mẫu | 4.0 | 1.4.1 | TU |
| G3.1 | <i>Lựa chọn</i> được tiêu chí và phương pháp để đánh giá | 3.5 | 2.2.3 | TU |
| G3.2 | <i>Xây dựng</i> được các kịch bản thử nghiệm | 3.5 | 2.2.3 | TU |
| G4.1 | <i>Lựa chọn</i> được các yếu tố liên quan đến quá trình mô hình hóa hệ thống | 4.0 | 4.3.3 | T |
| G4.2 | <i>Đánh giá</i> được hiệu năng của các mô hình phân lớp, phân cụm | 4.0 | 4.3.3 | TU |

5. Đánh giá học phần

| Thành phần đánh giá (1) | Bài đánh giá (2) | CDR học phần (G _{x.x}) (3) | Tỷ lệ (%) (4) |
|--|---|---|---------------|
| A1. Ý thức học tập và Hồ sơ học phần | | | 10% |
| A1.1. Đánh giá ý thức, thái độ học tập | | | |
| | Đánh giá sự chuyên cần (theo quy định của trường), thái độ | G1.2- G1.3 G2.1-G2.2 | 10% |
| A1.2. Hồ sơ học phần | | | 20% |
| | - Đánh giá 01 Project cá nhân qua hệ thống LMS Lưu ý: Giảng viên dạy lý thuyết thực hiện đánh giá và lưu hồ sơ | G1.2, G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | 20% |
| A1.3. Đánh giá định kỳ | | | 20% |
| | Bài kiểm tra trắc nghiệm <i>Ghi chú:</i> Trung tâm ĐBCL tổ chức thi | G1.1, G1.2, G1.3, G2.1 | 20% |
| A2. Đánh giá thực hành và thi kết thúc học phần | | | 50% |
| A2.1. Đánh giá thực hành | | | |
| | Đánh giá 01 bài thi thực hành <i>Ghi chú:</i> Giảng viên dạy thực hành đánh giá và lưu lại hồ sơ | G1.3, G2.2, G3.1, G3.2, G4.1, G4.2 | 1/3 × 50% |
| A2.2. Thi kết thúc học phần | | | |
| | Thi tự luận (Thời gian làm bài 90 phút) | G1.3, G2.1, G2.2, G4.1, G4.2 | 2/3 × 50% |
| Công thức: $A1.1 \times 0.1 + A1.2 \times 0.2 + A1.3 \times 0.2 + (A2.1 + A2.2 \times 2)/3 \times 0.5$ | | | |

6. Nội dung và kế hoạch giảng dạy

6.1. Nội dung giảng dạy

6.1.1. Lý thuyết

| Nội dung | CDR học phần | Bài đánh giá |
|--|---------------|------------------------|
| Chương 1. Tổng quan về khai phá dữ liệu 1.1. Các khái niệm, thuật ngữ 1.1.1. Dữ liệu, thông tin và tri thức 1.1.2. Khái niệm khai phá dữ liệu 1.2. Quá trình khai phá dữ liệu | G1.1, G1.2 | A1.1, A1.2, A1.3 |

| | | |
|--|---|--|
| 1.2.1. Quá trình khai phá dữ liệu 1.2.2. Dữ liệu và tiền xử lý dữ liệu 1.2.3. Các tác vụ/chức năng trong khai phá dữ liệu 1.3. Liên hệ giữa khai phá dữ liệu với các ngành liên quan 1.3.1. Khai phá dữ liệu với cơ sở dữ liệu 1.3.2. Khai phá dữ liệu với lý thuyết thống kê 1.3.3. Khai phá dữ liệu với học máy 1.4. Một số thuật toán tiêu biểu trong khai phá dữ liệu 1.5. Ứng dụng của khai phá dữ liệu 1.6. Một số công cụ thực hiện khai phá dữ liệu | | |
| Chương 2. Dữ liệu và tiền xử lý dữ liệu 2.1. Dữ liệu 2.1.1. Các loại dữ liệu 2.1.2. Chất lượng dữ liệu 2.2. Thống kê mô tả 2.3. Tiền xử lý dữ liệu 2.3.1. Kỹ thuật làm sạch dữ liệu 2.3.3. Kỹ thuật tích hợp và chuyển đổi dữ liệu 2.3.3. Kỹ thuật lấy mẫu dữ liệu 2.3.4. Kỹ thuật trích chọn đặc trưng 2.4. Bài tập | G1.3 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| Chương 3. Phân lớp dữ liệu 3.1. Bài toán phân lớp và dự báo 3.2. Tiếp cận chung giải bài toán phân lớp 3.2.1. Xây dựng mô hình phân lớp 3.2.2. Sử dụng mô hình phân lớp 3.3. Phân lớp cây quyết định 3.3.1. Cây quyết định 3.3.2. Xây dựng cây quyết định 3.3.3. Thuật toán Hunt 3.3.4. Ví dụ về phân lớp cây quyết định 3.3.5. Một số độ đo cho lựa chọn sự phân tách 3.3.5. Thuật toán phân lớp ID3, C45 3.3.6. Điểm mạnh và điểm yếu của phân lớp cây quyết định 3.4. Một số thuật toán phân lớp tiêu biểu 3.4.1. Naïve Bayes 3.4.2. KNN 3.6. Đánh giá mô hình phân lớp 3.6.1. Các tiêu chí đánh giá 3.6.2. Các chỉ số đánh giá 3.6.3. Phương pháp đánh giá 3.6.4. So sánh hai bộ phân lớp trên cùng 1 tập dữ liệu | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| Chương 4. Phân cụm dữ liệu 4.1. Tổng quan về phân cụm dữ liệu | G1.3, G2.1, G2.2, G3.1, | A1.1, |

| | | |
|---|------------------|------------------------------|
| 4.2. Các bước thực hiện phân cụm 4.3. Độ đo khoảng cách, độ đo tương tự 4.4. Phân cụm K-means 4.4.1. Thuật toán K-means 4.4.2. K-Means cải tiến 4.4.3. Điểm mạnh và điểm yếu của K-means 4.5. Phân cụm phân cấp 4.4.1. Các tiếp cận phân cụm phân cấp 4.4.2. Sự tương tự giữa các cụm 4.4.3. Điểm mạnh và điểm yếu của phân cụm phân cấp 4.6. Đánh giá mô hình phân cụm 4.6.1. Tổng quan 4.4.2. Độ cố kết và độ phân tách 4.4.3. Xác định số cụm tối ưu | G3.2, G4.1, G4.2 | A2.1, A2.2 |
| Chương 5. Khai phá luật kết hợp 5.1. Bài toán khai phá luật kết hợp và ứng dụng 5.2. Các khái niệm cơ bản 5.2.1. Cơ sở dữ liệu giao dịch 5.2.2. Luật kết hợp, độ hỗ trợ và độ tin cậy của luật 5.2.3. Độ hỗ trợ của tập mục 5.3. Tập ứng viên và tập thường xuyên 5.4 Tạo luật kết hợp từ tập mục thường xuyên 5.5. Một số thuật toán cải tiến của Apriori 5.6. Các thách thức và ứng dụng của khai phá luật kết hợp | G2.1, G2.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |

6.1.2. Dạy thực hành

| Buổi thực hành (mỗi buổi 2 tiết) | Nội dung | CĐR học phần | Bài đánh giá |
|-------------------------------------|--|--|------------------------------|
| 1 | Bài 1: Làm quen với công cụ và các ví dụ về khai phá dữ liệu | G1.1, G1.2 | A1.1, A1.3 |
| 2 | Bài 2: Tiền xử lý dữ liệu | G1.3, G4.1 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 3 - 6 | Bài 4: Phân lớp và dự báo | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 7 - 9 | Bài 5: Phân cụm dữ liệu | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 10 - 11 | Bài 5: Tìm tập mục thường xuyên và khai phá luật kết hợp | G2.1, G2.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |

| | | | |
|----|--------------------|--|-------|
| 12 | Kiểm tra thực hành | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A 2.1 |
|----|--------------------|--|-------|

6.2 Kế hoạch dạy học

Lý thuyết: (Mỗi tuần 2 tiết lý thuyết)

| Tuần (1) | Nội dung (2) | Hình thức tổ chức DH (3) | Chuẩn bị của SV (4) | CĐR học phần (5) | Bài đánh giá (6) |
|----------|---|--|---|------------------|------------------------------|
| 1 | Chương 1. Tổng quan về khai phá dữ liệu | <ul style="list-style-type: none"> - Giới thiệu đề cương chi tiết, tài liệu học tập, các yêu cầu của học phần, các hình thức đánh giá - Kết hợp giảng và trình chiếu slide - Minh họa các hệ thống khai phá dữ liệu | <ul style="list-style-type: none"> - Tài liệu phục vụ học tập - Đọc chương 1 tài liệu [1] | G1.1, G1.2 | A1.1, A1.2, A1.3 |
| 2 | Chương 2. Dữ liệu và tiền xử lý dữ liệu 2.1. Dữ liệu 2.1.1. Các loại dữ liệu 2.1.2. Chất lượng dữ liệu 2.2. Thống kê mô tả | <ul style="list-style-type: none"> - Giảng lý thuyết: kết hợp giảng và trình chiếu slide, thuyết trình và hỏi đáp - Minh họa các loại dữ liệu - Triển khai bài tập cá nhân | <ul style="list-style-type: none"> - Tài liệu phục vụ học tập - Tự học: Triển khai python, IDE và các thư viện liên quan đến khai phá dữ liệu - Đọc chương 2 tài liệu [2], Mục 2.1, 2.2 chương 2 [2]. Mục 3.2 chương 3 [2] | G1.3 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 3 | Chương 2. Dữ liệu và tiền xử lý dữ liệu (tiếp) 2.3. Tiền xử lý dữ liệu 2.3.1. Kỹ thuật làm sạch dữ liệu 2.3.2. Kỹ thuật tích hợp và chuyển đổi dữ liệu | <ul style="list-style-type: none"> - Kiểm tra nhanh các câu trắc nghiệm kiến thức chương 1 - Kết hợp giảng và trình chiếu slide; - Minh họa tiền xử lý dữ liệu - Phân tích ý nghĩa của tiền xử lý dữ liệu | <ul style="list-style-type: none"> - Tài liệu phục vụ học tập - Tài các bộ dữ liệu mẫu và tìm hiểu đặc điểm - Đọc chương 2 tài liệu [1] | G1.3 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 4 | Chương 2. Dữ liệu và tiền xử lý dữ liệu (tiếp) | <ul style="list-style-type: none"> - Kết hợp giảng và trình chiếu slide; | <ul style="list-style-type: none"> - Tài liệu phục vụ học tập - Tự học: Triển khai ví dụ đơn | G1.3 | A1.1, A1.2, A1.3, |

| | | | | | |
|----------|--|--|---|---|---|
| | <p>2.3.3. Kỹ thuật lấy mẫu dữ liệu</p> <p>2.3.4. Kỹ thuật trích chọn đặc trưng</p> <p>2.4. Bài tập</p> | <p>- Phân tích và minh họa lợi ích của lấy mẫu và trích chọn đặc trưng</p> | <p>giảm ở chương 1 quyền [3]</p> <p>- Đọc chương 2 tài liệu [1]</p> | | <p>A2.1,</p> <p>A2.2</p> |
| 5 | <p>Chương 3. Phân lớp dữ liệu</p> <p>3.1. Bài toán phân lớp và dự báo</p> <p>3.2. Tiếp cận chung giải bài toán phân lớp</p> <p>3.2.1. Xây dựng mô hình phân lớp</p> <p>3.2.2. Sử dụng mô hình phân lớp</p> <p>3.3. Phân lớp cây quyết định</p> <p>3.3.1. Cây quyết định</p> <p>3.3.2. Xây dựng cây quyết định</p> <p>3.3.3. Thuật toán Hunt</p> <p>3.3.4. Ví dụ về phân lớp cây quyết định</p> | <p>- Kiểm tra trắc nghiệm câu hỏi liên quan kiến thức chương 2</p> <p>- Giáo viên nêu vấn đề, thảo luận để giải quyết vấn đề</p> <p>- Kết hợp giảng và trình chiếu slide</p> <p>- Lấy ví dụ thực tế về ứng dụng của phân lớp</p> | <p>- Tài liệu phục vụ học tập</p> <p>- Đọc chương 4 tài liệu [1], Mục 4.3 chương 4 tài liệu [2]</p> <p>- Tự học: + Nghiên cứu ví dụ về dự báo chơi tennis trong chương 3 [4] + Tiếp tục bài tập cá nhân</p> | <p>G1.3,</p> <p>G2.1,</p> <p>G2.2,</p> <p>G3.1,</p> <p>G3.2,</p> <p>G4.1,</p> <p>G4.2</p> | <p>A1.1,</p> <p>A1.2,</p> <p>A1.3,</p> <p>A2.1,</p> <p>A2.2</p> |
| 6 | <p>Chương 3. Phân lớp dữ liệu (tiếp)</p> <p>3.3.5. Một số độ đo cho lựa chọn sự phân tách</p> <p>3.3.5. Thuật toán phân lớp ID3, C45</p> <p>3.3.6. Điểm mạnh và điểm yếu của phân lớp cây quyết định</p> | <p>- Cho SV trao đổi về các cây khác nhau được xây dựng từ 1 tập dữ liệu</p> <p>- Kết hợp giảng và trình chiếu slide</p> <p>- Phân biệt ID3 với C45</p> | <p>- Tài liệu phục vụ học tập</p> <p>- Tự học: Làm bài tập 1, 2 và 3 mục 4.8 chương 4 tài liệu [2]</p> | <p>G1.3,</p> <p>G2.1,</p> <p>G2.2,</p> <p>G3.1,</p> <p>G3.2,</p> <p>G4.1,</p> <p>G4.2</p> | <p>A1.1,</p> <p>A1.2,</p> <p>A1.3,</p> <p>A2.1,</p> <p>A2.2</p> |
| 7 | <p>Chương 3. Phân lớp dữ liệu (tiếp)</p> <p>3.4. Một số thuật toán phân lớp tiêu biểu</p> <p>3.4.1. Naïve Bayes</p> <p>3.4.2. Thuật toán KNN</p> | <p>- Kết hợp giảng và trình chiếu slide</p> <p>- Phân tích ý tưởng, cách tiếp cận và ưu nhược điểm, của thuật toán Naïve Bayes, KNN</p> | <p>- Tài liệu phục vụ học tập</p> <p>- Đọc mục 5.2 và 5.3 Chương 5 tài liệu [2]</p> <p>- Tự học: Làm bài tập 7 mục 5.10 chương 5 [2]</p> | <p>G1.3,</p> <p>G2.1,</p> <p>G2.2,</p> <p>G3.1,</p> <p>G3.2,</p> <p>G4.1,</p> <p>G4.2</p> | <p>A1.1,</p> <p>A1.2,</p> <p>A1.3,</p> <p>A2.1,</p> <p>A2.2</p> |
| 8 | <p>Chương 3. Phân lớp dữ liệu</p> <p>3.6. Đánh giá mô hình phân lớp</p> <p>3.6.1. Các tiêu chí đánh giá</p> <p>3.6.2. Các chỉ số đánh giá</p> <p>3.6.3. Phương pháp đánh giá</p> <p>3.6.4. So sánh hai bộ phân lớp trên cùng một tập dữ liệu</p> | <p>- Kiểm tra trắc nghiệm kiến thức về phân lớp</p> <p>- Kết hợp giảng và trình chiếu slide</p> <p>- Cho sinh viên thảo luận về ý nghĩa của mỗi chỉ</p> | <p>- Tài liệu phục vụ học tập</p> <p>- Tự học: Làm bài tập 11 mục 4.8 chương 4 tài liệu [2]</p> | <p>G1.3,</p> <p>G2.1,</p> <p>G2.2,</p> <p>G3.1,</p> <p>G3.2,</p> <p>G4.1,</p> <p>G4.2</p> | <p>A1.1,</p> <p>A1.2,</p> <p>A1.3,</p> <p>A2.1,</p> <p>A2.2</p> |

| | | số, mỗi phương pháp đánh giá | | | |
|-----------|---|--|--|--|------------------------------|
| 9 | Chương 4. Phân cụm dữ liệu 4.1. Tổng quan về phân cụm dữ liệu 4.2. Các bước thực hiện phân cụm 4.3. Độ đo khoảng cách, độ đo tương tự | - Kết hợp giảng và trình chiếu slide - Minh họa các hệ thống phân cụm - Cho sinh viên thảo luận về phân lớp và phân cụm | - Tài liệu phục vụ học tập - Tự học: - Đọc chương 7 tài liệu [1], Chương 9 tài liệu [2] | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A2.1, A2.2 |
| 10 | Chương 4. Phân cụm dữ liệu (tiếp) 4.4. Phân cụm K-means 4.4.1. Thuật toán K-means 4.4.2. K-Means cải tiến 4.4.3. Điểm mạnh và điểm yếu của K-means | - Kiểm tra trắc nghiệm kiến thức - Kết hợp giảng và trình chiếu slide - Làm bài tập | - Bản cứng slide bài giảng phục vụ học phần. - Đọc chương 5 tài liệu [2] | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A2.1, A2.2 |
| 11 | Chương 4. Phân cụm dữ liệu (tiếp) 4.5. Phân cụm phân cấp 4.5.1. Các tiếp cận phân cụm phân cấp 4.5.2. Sự tương tự giữa các cụm 4.5.3. Điểm mạnh và điểm yếu của phân cụm phân cấp | - Kết hợp giảng, trình chiếu slide và đặt câu hỏi - Minh họa hệ thống phân cụm phân cấp | - Tài liệu phục vụ học tập - Hoàn thành bài tập cá nhân, nạp vào LMS - Đọc chương 5 tài liệu [1] | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A2.1, A2.2 |
| 12 | Chương 4. Phân cụm dữ liệu (tiếp) 4.6. Đánh giá mô hình phân cụm 4.6.1. Tổng quan 4.5.2. Độ cố kết và độ phân tách 4.5.3. Xác định số cụm tối ưu | - Kết hợp giảng, trình chiếu slide và đặt câu hỏi - Kiểm tra trắc nghiệm kiến thức về phân cụm | - Tài liệu phục vụ học tập - Thực hành - Đọc chương 6 tài liệu [2] - Nạp bài tập cá nhân qua hệ thống LMS | G1.3, G2.1, G2.2, G3.1, G3.2, G4.1, G4.2 | A1.1, A2.1, A2.2 |
| 13 | Chương 5. Khai phá luật kết hợp 5.1. Bài toán khai phá luật kết hợp và ứng dụng 5.2. Các khái niệm cơ bản 5.2.1. Cơ sở dữ liệu giao dịch 5.2.2. Luật kết hợp, độ hỗ trợ và độ tin cậy của luật 5.2.3. Độ hỗ trợ của tập mục 5.3. Tập ứng viên và tập thường xuyên | - GV kết hợp giảng và trình chiếu slide; - Minh họa ví dụ/hệ thống khai phá luật kết hợp - Cho sinh viên lấy các ví dụ khác về khai phá luật kết hợp trong thực tế | - Tài liệu phục vụ học tập - Tự học: Tìm hiểu hệ thống gợi ý phim trong chương 4 [3] - Đọc chương 4 tài liệu [1] | G2.1, G2.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |
| 14 | Chương 5. Khai phá luật kết hợp (tiếp) | - Kết hợp giảng và trình chiếu slide | - Tài liệu phục vụ học tập | G2.1, G2.2 | A1.1, A1.2, A1.3, |

| | | | | | |
|-----------|--|--|--|------------|------------------------------|
| | 5.4 Tạo luật kết hợp từ tập mục thường xuyên 5.5. Một số thuật toán cải tiến của Apriori | - Kiểm tra trắc nghiệm kiến thức về khai phá luật kết hợp | - Tự học: Làm bài tập 1, 2 và 6 mục 6.10 chương 6 tài liệu [2] - Đọc chương 4 tài liệu [1]. | | A2.1, A2.2 |
| 15 | Chương 5. Khai phá luật kết hợp (tiếp) 5.6. Các thách thức và ứng dụng của khai phá luật kết hợp Tổng kết học phần | - Kết hợp giảng và trình chiếu slide - Giải đáp thắc mắc - Thuyết giảng về các nội dung đã học | - Tài liệu phục vụ học tập - Tự học: Đọc chương 4 tài liệu [1]. | G2.1, G2.2 | A1.1, A1.2, A1.3, A2.1, A2.2 |

Thực hành: 12 buổi (mỗi tuần 2 tiết, bắt đầu từ tuần 3)

| Buổi thực hành (2 tiết) | Nội dung | Hình thức tổ chức dạy học | Chuẩn bị của sinh viên | CDR học phần | Bài đánh giá |
|--------------------------------|---|---|--|------------------------|---------------------|
| 1 | Bài 1: Làm quen với công cụ nguồn mở cho khai phá dữ liệu | Quản triệ nội quy, quy định học thực hành; Cài đặt python, IDE, các thư viện thường dùng; Thực hiện chạy thử 1 số ví dụ đơn giản trong bài thực hành 1 và chương 1 quyển [3] | Chuẩn bị nội dung theo yêu cầu của Bài thực hành 1 | G1.1 | |
| 2 - 3 | Bài 2: Tiền xử lý dữ liệu | Xử lý dữ liệu bị thiếu giá trị; Xử lý giá trị ngoại lai Giảm chiều và rời rạc hóa dữ liệu | Chuẩn bị nội dung theo yêu cầu của Bài thực hành 2 SV tự thực hành theo nội dung; Nạp bài vào hệ thống LMS | G2.2, G3.1 | A1.3 A2.1 |
| 4 – 7 | Bài 4: Phân lớp và dự báo | Áp dụng thuật toán ID3, C45, Naïve Bayes phân lớp dữ liệu trên tập dữ liệu mẫu; Đánh giá hiệu năng phân lớp; Tổ chức cho SV thảo luận, nhận xét và góp ý; GV tổng hợp, nhận xét về ưu điểm, nhược điểm của các giải pháp Kiểm tra thực hành | Chuẩn bị nội dung theo yêu cầu của Bài thực hành 4 - Làm bài thi thực hành 60 phút | G2.1, G2.2, G3.1, G3.2 | A1.3 A2.1, A2.2 |

| | | | | | |
|----------------|--|--|---|------------------------------|------------------|
| 8 - 10 | Bài 5: Phân cụm dữ liệu | Áp dụng thuật toán kmeans và phân cụm phân cấp; Đánh giá kết quả phân cụm | Chuẩn bị nội dung theo yêu cầu của Bài thực hành 5 | G2.1, G4.1, G4.2, G5.1, G5.2 | A1.3, A2.1, A2.2 |
| 11 - 12 | Bài 3: Tìm tập mục thường xuyên và khai phá luật kết hợp | SV thực hành theo nội dung bài thực hành; GV theo dõi và nhận xét về buổi thực hành | - Chuẩn bị nội dung theo yêu cầu của Bài thực hành 3; - Nạp bài vào hệ thống LMS | G2.1, G3.2 | A1.3, A2.1, A2.2 |

7. Nguồn học liệu

Giáo trình:

- [1] Hà Quang Thụy và cộng sự, *Web Data mining*, NXB Giáo dục Việt Nam, 2011
- [2] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, *Introduction to Data Mining*, Pearson, 2013

Tài liệu tham khảo:

- [3] Wes McKinney, *Python for Data Analysis*, 2nd Edition, O'Reilly, 2017
- [4] Robert Layton, *Learning Data Mining with Python*, 2nd Edition, Packt Publishing, 2017

8. Quy định của học phần

- ✓ Dự lớp theo đúng quy chế
- ✓ Thực hiện đầy đủ các bài tập, tự học theo yêu cầu của giảng viên
- ✓ Thực hiện đầy đủ các bài đánh giá

9. Phụ trách học phần

- Viện Kỹ thuật và Công nghệ - Bộ môn Khoa học máy tính và Công nghệ phần mềm
- Địa chỉ: Tầng 1 - Nhà A0 - Đại học Vinh
- Email: yktcn.khmt_cnpm@vinhuni.edu.vn