

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN

1. Tên học phần: Khai khoáng dữ liệu (Data mining)

- Mã số học phần: CT312
- Số tín chỉ học phần: 3 tín chỉ
- Số tiết học phần: 30 tiết lý thuyết, 30 tiết thực hành và 75 tiết tự học.

2. Đơn vị phụ trách học phần:

- Bộ môn: Khoa học máy tính
- Khoa/Viện/Trung tâm/Bộ môn: Khoa Công nghệ thông tin & Truyền thông

3. Điều kiện:

- Điều kiện tiên quyết: CT202 (Nguyên lý máy học)
- Điều kiện song hành: không có

4. Mục tiêu của học phần:

Mục tiêu	Nội dung mục tiêu	CĐR CTĐT
4.1	<ul style="list-style-type: none"> - Trình bày được quy trình khai khoáng dữ liệu và các lĩnh vực ứng dụng. - Vận dụng được các kỹ thuật tiền xử lý dữ liệu trong quá trình khai khoáng phù hợp với dữ liệu thực tiễn - Vận dụng các giải thuật máy học, luật kết hợp, phương pháp tập hợp mô hình để rút trích được các tri thức hữu ích trong quá trình khai phá dữ liệu nhằm giải quyết vấn đề thực tế. - Nhận diện được các bài toán và các kỹ thuật khai phá dữ liệu phù hợp bài toán: giải thuật cây quyết định, rừng ngẫu nhiên, gom cụm, luật kết hợp.... 	2.1.3.c
4.2	<ul style="list-style-type: none"> - Phân tích, thiết kế và xây dựng các hệ thống khai khoáng dữ liệu. - Vận dụng được các giải thuật cơ bản và từ đó đề xuất các giải pháp khai phá dữ liệu cho các ứng dụng khác nhau. - Đánh giá hiệu quả của các thuật toán khai phá dữ liệu khác nhau. 	2.2.1.b
4.3	Kỹ năng làm việc nhóm, thuyết trình, phản biện và lập kế hoạch	2.2.2. a, 2.2.2. b 2.3 a

5. Chuẩn đầu ra của học phần:

CDR HP	Nội dung chuẩn đầu ra	Mục tiêu	CDR CTĐT
	Kiến thức		
CO1	Mô tả các khái niệm cơ bản và ứng dụng của khai phá dữ liệu	4.1	2.1.3.c
CO2	Trình bày các bước trong quá trình khai phá dữ liệu để rút trích được tri thức hữu ích	4.1	2.1.3.c
CO3	Phân biệt và vận dụng phù hợp các kỹ thuật tiền xử lý dữ liệu cho các loại dữ liệu và bài toán khác nhau.	4.1	2.1.3.c
CO4	Trình bày được ý tưởng và trường hợp sử dụng của giải thuật khai phá dữ liệu như hồi quy, phân lớp, gom cụm và luật kết hợp	4.1	2.1.3.c
	Kỹ năng		
CO5	Phân tích và vận dụng các kỹ thuật phù hợp của khai khoáng dữ liệu trong giải quyết vấn đề cụ thể	4.2	2.2.1.b
CO6	Khả năng lập kế hoạch, làm việc nhóm, thuyết trình, phản biện	4.3	2.1.3.b 2.1.3.c 2.2.1.b 2.2.2. b
	Thái độ/Mức độ tự chủ và trách nhiệm		
CO7	Hoàn thành bài tập nhóm đúng thời hạn và yêu cầu, trung thực và nghiêm túc trong quá trình đánh giá	4.3	2.3 a, 2.3 b

6. Mô tả tóm tắt nội dung học phần:

Học phần khai khoáng dữ liệu cung cấp kiến thức về quy trình khai khoáng dữ liệu, các phương pháp biến đổi dữ liệu thô đã thu thập thành thông tin và kiến thức hữu ích, tri thức tiềm, thú. Vận dụng các giải thuật khai khoáng dữ liệu phổ biến như: K láng giềng, Bayes thơ ngây, cây quyết định, rừng ngẫu nhiên, phương pháp tập hợp mô hình, giải thuật gom cụm, luật kết hợp để giải quyết các vấn đề cụ thể trong thực tế.

7. Cấu trúc nội dung học phần:

7.1. Lý thuyết

	Nội dung	Số tiết	CDR HP
Chương 1.	Giới thiệu về khám phá tri thức và khai khoáng dữ liệu	2	
1.1.	Khám phá tri thức và khai khoáng dữ liệu	0.5	CO1
1.2.	Các lĩnh vực ứng dụng	0.5	CO1
1.3.	Quá trình khám phá tri thức	0.5	CO1, CO2
1.4.	Xu hướng phát triển	0.5	CO1
Chương 2.	Tiền xử lý dữ liệu	4	
2.1.	Làm sạch dữ liệu	0.5	CO2, CO3
2.2.	Tích hợp dữ liệu	0.5	CO2, CO3

	Nội dung	Số tiết	CDR HP
2.3.	Thu gọn dữ liệu	0.5	CO2, CO3
2.4.	Hiển thị dữ liệu	0.5	CO2, CO3
	Bài tập	2	CO2, CO3
Chương 3.	Phương pháp hồi quy	4	
3.1.	Giới thiệu bài toán hồi quy	0.5	CO2, CO4
3.2.	Hồi quy tuyến tính	0.5	CO2, CO4
3.3.	Hồi quy phi tuyến	1	CO2, CO4
	Bài tập	2	CO2, CO4
Chương 4.	Phân lớp dữ liệu	7	
4.1.	Tổng quan về phân lớp dữ liệu	0.5	CO2, CO4
4.2.	Phân lớp dữ liệu với cây quyết định, rừng ngẫu nhiên	1	CO2, CO4
4.3.	Phân lớp dữ liệu với giải thuật bayes thơ ngây	1	CO2, CO4
4.4.	Phân lớp dữ liệu với mạng Neural	1	CO2, CO4
4.5.	Phương pháp tập hợp mô hình	1.5	CO2, CO4
	Bài tập	2	CO2, CO4
Chương 5.	Gom cụm dữ liệu	7	
5.1.	Tổng quan về gom cụm dữ liệu	0.5	CO2, CO4
5.2.	Gom cụm dữ liệu bằng phân hoạch	1	CO2, CO4
5.3.	Gom cụm dữ liệu bằng phân cấp	1	CO2, CO4
5.4.	Gom cụm dữ liệu dựa trên mật độ	1	CO2, CO4
5.5.	Các phương pháp gom cụm dữ liệu khác	1.5	CO2, CO4
	Bài tập	2	CO2, CO4
Chương 6.	Khai phá luật kết hợp	6	
6.1.	Khai phá mẫu phổ biến và luật kết hợp	1	CO2, CO4
6.2.	Giải thuật Apriori	2	CO2, CO4
6.3.	Đánh giá luật kết hợp	1	CO2, CO4
	Bài tập	2	CO2, CO4

7.2. Thực hành

	Nội dung	Số tiết	CDR HP
Bài 1.	Phương pháp hồi quy + tiền xử lý dữ liệu	5	CO1, CO2, CO3
Bài 2.	Phân lớp dữ liệu + tiền xử lý dữ liệu	5	CO2, CO4
Bài 3.	Phân lớp dữ liệu (tt) + tiền xử lý dữ liệu	5	CO2, CO4
Bài 4.	Gom cụm dữ liệu + tiền xử lý dữ liệu	5	CO2, CO4
Bài 5.	Khai phá luật kết hợp + tiền xử lý dữ liệu	5	CO2, CO4
Bài 6.	Bài thực hành tổng hợp	5	CO2, CO3, CO4, CO5

8. Phương pháp giảng dạy:

- Giờ lý thuyết:
 - Giảng viên trình bày lý thuyết.
 - Giảng viên đặt vấn đề mới bằng các câu hỏi để kích thích tư duy sáng tạo của sinh viên.

- Giảng viên ôn tập bài cũ thông qua các câu hỏi để giúp sinh viên nhớ lâu, nhớ sâu, tăng tính tự học, chủ động của sinh viên.
- Giờ bài tập:
 - Sinh viên làm bài tập và đại diện sinh viên lên bảng trình bày lời giải.
 - Giảng viên phân tích và sửa các lời giải, giảng viên phân tích những điểm đúng/sai.
 - Giảng viên hướng dẫn, gợi ý các bước hoàn thành bài tập khó.

9. Nhiệm vụ của sinh viên:

Sinh viên phải thực hiện các nhiệm vụ như sau:

- Tham gia trên 80% giờ học lý thuyết trên lớp.
- Thực hiện đầy đủ các bài tập trên lớp và bài tập nhóm.
- Tham gia kiểm tra giữa kỳ.
- Tham gia thi kết thúc học phần.
- Chủ động tổ chức thực hiện giờ tự học.

10. Đánh giá kết quả học tập của sinh viên:

10.1. Cách đánh giá

Sinh viên được đánh giá tích lũy học phần như sau:

TT	Điểm thành phần	Quy định	Trọng số	CDR HP
1	Điểm chuyên cần	Tham gia làm bài tập, kiểm tra tại lớp	10%	CO7
2	Điểm giữa kỳ	Báo cáo bài tập nhóm: lập trình, làm slide báo cáo và thuyết trình	30%	CO1, CO2, CO3, CO4, CO5, CO6, CO7
3	Điểm cuối kỳ	Thi tự luận	60%	CO1, CO2, CO3, CO4, CO5, CO7

10.2. Cách tính điểm

- Điểm đánh giá thành phần và điểm thi kết thúc học phần được chấm theo thang điểm 10 (từ 0 đến 10), làm tròn đến một chữ số thập phân.
- Điểm học phần là tổng điểm của tất cả các điểm đánh giá thành phần của học phần nhân với trọng số tương ứng. Điểm học phần theo thang điểm 10 làm tròn đến một chữ số thập phân, sau đó được quy đổi sang điểm chữ và điểm số theo thang điểm 4 theo quy định về công tác học vụ của Trường.

11. Tài liệu học tập:

Thông tin về tài liệu	Số đăng ký cá biệt
[1] Ian H. Witten, Eibe Frank, Data mining: practical machine learning tools and techniques, 2nd edition, Morgan Kaufmann Publishers, 2005	CNTT.001068
[2] Giáo trình các hệ tri thức và khai thác dữ liệu Đỗ Thanh Nghị, Lê Thanh Vân, Nxb. Đại học Cần Thơ, 2012	CNTT.002816, CNTT.002817, CNTT.002818

12. Hướng dẫn sinh viên tự học:

Tuần	Nội dung	Lý thuyết (tiết)	Thực hành (tiết)	Nhiệm vụ của sinh viên
1	Chương 1: Giới thiệu về khám phá tri thức và khai khoáng dữ liệu 1.1. Khám phá tri thức và khai khoáng dữ liệu. 1.2. Các lĩnh vực ứng dụng. 1.3. Quá trình khám phá tri thức. 1.4. Xu hướng phát triển.	2	0	- Nghiên cứu trước: + Tài liệu [1]: Chương 1
2	Chương 2: Chuẩn bị dữ liệu 2.1. Làm sạch dữ liệu. 2.2. Tích hợp dữ liệu. 2.3. Thu gọn dữ liệu. 2.4. Hiển thị dữ liệu.	4	0	- Nghiên cứu trước: + Tài liệu [1,2]: Chương 2, 4
3	Chương 3: Các phương pháp dự báo 3.1. Giới thiệu dự báo. 3.2. Hồi quy tuyến tính. 3.3. Hồi quy phi tuyến.	4	0	- Nghiên cứu trước: + Tài liệu [2]: Chương 8, 9
4	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, rừng ngẫu nhiên 4.3. Phân lớp dữ liệu với bayes thơ ngây 4.4. Phân lớp dữ liệu với mạng Neural. 4.5. Phương pháp tập hợp mô hình	7	0	- Nghiên cứu trước: + Tài liệu [1]: Chương 11, 12, 14 + Tài liệu [2]: các giải thuật phân lớp
5	Chương 5: Góm cụm dữ liệu 5.1. Tổng quan về gom cụm dữ liệu. 5.2. Góm cụm dữ liệu bằng phân hoạch. 5.3. Góm cụm dữ liệu bằng phân cấp. 5.4. Góm cụm dữ liệu dựa	7	0	- Nghiên cứu trước: + Tài liệu [1]: Chương 19 + Tài liệu [2]: các giải thuật gom cụm

Tuần	Nội dung	Lý thuyết (tiết)	Thực hành (tiết)	Nhiệm vụ của sinh viên
	trên mặt độ. 5.5. Các phương pháp gom cụm dữ liệu khác.			
6	Chương 6: Khai phá luật kết hợp 6.1. Các giai đoạn của quá trình tìm luật kết hợp. 6.2. Những hướng tiếp cận chính.	6	0	- Nghiên cứu trước: + Tài liệu [1]: Chương 23 + Tài liệu [2]: luật kết hợp
7	Bài thực hành 2: Phương pháp dự báo	0	5	- Nghiên cứu trước: + Tài liệu [1]: Chương 8, 9 + Tài liệu [2]: Chương 1, 2, 4
8	Bài thực hành 3: Phân lớp dữ liệu.	0	5	- Nghiên cứu trước: + Tài liệu [1]: Chương 11, 12, 14 + Tài liệu [2]: Chương 1, 2, 4
9	Bài thực hành 3: Phân lớp dữ liệu (tt)	0	5	- Nghiên cứu trước: + Tài liệu [1]: Chương 11, 12 + Tài liệu [2]: Chương 1, 2, 4
10	Bài thực hành 4: Gom cụm dữ liệu.	0	5	- Nghiên cứu trước: + Tài liệu [1]: Chương 19 + Tài liệu [2]: Chương 1, 2, 4
11	Bài thực hành 5: Khai phá luật kết hợp.	0	5	- Nghiên cứu trước: + Tài liệu [1]: Chương 23 + Tài liệu [2]: Chương 1, 2, 4
12	Bài thực hành 6: Bài thực hành tổng hợp.	0	5	- Tổng hợp các bài thực hành trước



TRẦN HIỆU TRƯỞNG
TRƯỞNG KHOA

Nguyễn Hữu Hòa

Cần Thơ, ngày 09 tháng 5 năm 2019

TRƯỞNG BỘ MÔN

Trần Nguyễn Minh Thư