

ĐỀ CƯƠNG MÔN HỌC

KHAI PHÁ DỮ LIỆU (Data mining)

Mã số MH : XXXXXXXXXX

- Số tín chỉ :	Tc (LT.BT&TH.TựHọc): 3	TCHP:	
- Số tiết :	Tổng: 60 LT:45 BT: TH: ĐA: BTL/TL:15		
- Đánh giá Thang điểm 10/10	Bài tập lớn Kiểm tra giữa kỳ Thuyết trình Thi cuối kỳ	30% 20% 50%	Làm bài tập theo nhóm Thuyết trình cá nhân Thi viết 90 phút
- Môn tiên quyết :	-		MS:
- Môn học trước :	-		MS:
- Môn song hành :	-		MS:
- CTĐT ngành	<i>Khoa học máy tính</i>		
Mã ngành :	60.48.01		
- Ghi chú khác :			

❖ Mục tiêu của môn học:

Khi hoàn tất môn học này, sinh viên có khả năng:

- Hiểu các khái niệm, công nghệ và việc sử dụng khai phá dữ liệu
- Hiểu cách khai thác dữ liệu trong việc khám phá tri thức
- Áp dụng công nghệ khai phá dữ liệu để hỗ trợ việc ra quyết định điều hành và mang tính chiến lược trong một tổ chức
- Đạt được kinh nghiệm thực tế với phần mềm khai phá dữ liệu

❖ Nội dung tóm tắt môn học:

Trong môn học này, tổng quan về khai phá dữ liệu trong việc hỗ trợ ra quyết định điều hành và mang tính chiến lược sẽ được trình bày. Các kỹ thuật khai phá dữ liệu được giới thiệu cho việc khám phá các thông tin ẩn từ dữ liệu thu thập được. Bên cạnh đó, môn học này cũng giới thiệu cách phân tích các nhu cầu kinh doanh cho việc khám phá tri thức để tạo nên lợi thế cạnh tranh và cách áp dụng các công nghệ khai phá dữ liệu một cách thích hợp để nhận dạng giá trị kinh doanh thực sự của các công nghệ khai phá dữ liệu này. Các đề tài cụ thể thuộc về môn học này bao gồm: các phương pháp và quá trình khai phá dữ liệu, các công nghệ khai phá dữ liệu, các ứng dụng và nghiên cứu cụ thể của khai phá dữ liệu.

❖ Tài liệu học tập

Giáo trình

- [1] Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, Second Edition, Morgan Kaufmann Publishers, 2006.
- [2] David Hand, Heikki Mannila, Padhraic Smyth, “Principles of Data Mining”, MIT Press, 2001

Sách tham khảo

- [3] David L. Olson, Dursun Delen, “Advanced Data Mining Techniques”, Springer-Verlag, 2008.
- [4] Graham J. Williams, Simeon J. Simoff, “Data Mining: Theory, Methodology, Techniques, and Applications”, Springer-Verlag, 2006.
- [5] ZhaoHui Tang, Jamie MacLennan, “Data Mining with SQL Server 2005”, Wiley Publishing, 2005.
- [6] Oracle, “Data Mining Concepts”, B28129-01, 2008.
- [7] Oracle, “Data Mining Application Developer’s Guide”, B28131-01, 2008.
- [8] The Internet

❖ Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học

Tri thức:

- Các khái niệm cơ bản về khai phá dữ liệu và việc hỗ trợ ra quyết định
- Các vấn đề phân tích và xử lý dữ liệu cho quá trình khai phá dữ liệu
- Các kỹ thuật khai phá dữ liệu
- Các ứng dụng khai phá dữ liệu

Kỹ năng nhận thức:

- Khả năng nhận dạng và hiểu ý nghĩa và vai trò của khai phá dữ liệu trong tình hình kinh tế-xã hội-khoa học-kỹ thuật ngày nay
- Khả năng nhận dạng và hiểu các vấn đề liên quan đến dữ liệu sẽ được khai phá và quá trình khai phá dữ liệu
- Khả năng nhận dạng và hiểu khả năng ứng dụng của khai phá dữ liệu vào các hoạt động cụ thể của một tổ chức

Kỹ năng thực hành môn học:

- Khả năng phân tích và xử lý dữ liệu cho quá trình khai phá dữ liệu
- Khả năng phát triển các kỹ thuật khai phá dữ liệu
- Khả năng phát triển ứng dụng khai phá dữ liệu
- Khả năng vận dụng các tiện ích hỗ trợ khai phá dữ liệu được cung cấp cùng với các hệ quản trị cơ sở dữ liệu phổ biến ngày nay như MS SQL Server và Oracle

Kỹ năng truyền đạt:

- Khả năng tham gia phân tích và xử lý dữ liệu cho quá trình khai phá dữ liệu
- Khả năng tham gia phát triển các kỹ thuật khai phá dữ liệu
- Khả năng tham gia phát triển ứng dụng khai phá dữ liệu

❖ Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Các khuyến cáo về trang bị phục vụ học tập và tài liệu tham khảo: sinh viên cài đặt hai DBMSs (MS SQL Server 2005/2008 và Oracle 10g/11g) và chuẩn bị tài liệu tham khảo cho việc tham dự giờ giảng trên lớp

Các yêu cầu đặc biệt khác:.

- Về thực hiện báo cáo tiểu luận: nhóm 2 người, bắt đầu thực hiện từ tuần 2 đến tuần 15, nộp đề cương của báo cáo tiểu luận vào tuần 4, nộp báo cáo tiểu luận vào tuần 13 và trình bày báo cáo tiểu luận vào tuần 14-15.

Cách tổng kết điểm: $20\% \text{điểm thi giữa kỳ} + 30\% \text{điểm báo cáo tiểu luận} + 50\% \text{điểm thi cuối kỳ}$ từ 5 trở lên (≥ 5.0) mới tính là đạt cả môn học.

- Đối với học viên là NCS, HV được coi là hoàn tất môn học nếu điểm môn học trên 7 điểm

❖ Nội dung chi tiết:

- **LÝ THUYẾT:** 45 tiết (3 tiết/tuần x 15 tuần)

Tuần	Nội dung	Tài liệu	Ghi chú
1	Chương 1: Tổng quan về khai phá dữ liệu <ul style="list-style-type: none"> 1.1. Khái niệm 1.2. Quá trình khai phá dữ liệu 1.3. Ý nghĩa 1.4. Vai trò 1.5. Ứng dụng 1.6. Tóm tắt 	[1-3, 8]	Hiểu Năm vững
2	Chương 2: Dữ liệu được khai phá <ul style="list-style-type: none"> 2.1. Kiểu dữ liệu 2.2. Các độ đo khoảng cách trong khai phá dữ liệu 2.3. Vấn đề thu thập và chuẩn bị dữ liệu 2.4. Vấn đề chất lượng dữ liệu 2.5. Trục quan hóa dữ liệu 2.6. Tóm tắt 	[1, 2, 8]	Hiểu, Năm vững, Tổng hợp
3,4	Chương 3: Các vấn đề tiền xử lý dữ liệu <ul style="list-style-type: none"> 3.1. Tổng quan về giai đoạn tiền xử lý dữ liệu 3.2. Tóm tắt hoá dữ liệu 3.3. Làm sạch dữ liệu 3.4. Tích hợp dữ liệu 3.5. Biến đổi dữ liệu 3.6. Thu giảm dữ liệu 3.7. Rời rạc hóa dữ liệu 3.8. Tạo cây phân cấp ý niệm 3.9. Biểu diễn dữ liệu 3.10. Tóm tắt 	[1,8]	Hiểu, Năm vững, Vận dụng, Tổng hợp
5,6	Chương 4: Phân loại dữ liệu <ul style="list-style-type: none"> 4.1. Tổng quan về phân loại dữ liệu 4.2. Phân loại dữ liệu với cây quyết định 4.3. Phân loại dữ liệu với mạng Bayesian 4.4. Phân loại dữ liệu với mạng Neural 4.5. Các phương pháp phân loại dữ liệu khác 4.6. Tóm tắt 	[1,3-6,8]	Hiểu, Năm vững, Vận dụng, Tổng hợp
7,8	Chương 5: Gom cụm dữ liệu <ul style="list-style-type: none"> 5.1. Tổng quan về gom cụm dữ liệu 5.2. Gom cụm dữ liệu bằng phân hoạch 	[1,3-6,8]	Hiểu, Năm vững,

Tuần	Nội dung	Tài liệu	Ghi chú
	5.3. Gom cụm dữ liệu bằng phân cấp 5.4. Gom cụm dữ liệu dựa trên mật độ 5.5. Gom cụm dữ liệu dựa trên mô hình 5.6. Các phương pháp gom cụm dữ liệu khác 5.7. Tóm tắt		Vận dụng, Tổng hợp
9,10	Chương 6: Luật kết hợp 6.1. Tổng quan về luật kết hợp 6.2. Biểu diễn luật kết hợp 6.3. Khám phá các mẫu thường xuyên 6.4. Khám phá các kết hợp với giải thuật Apriori và các biến thể của giải thuật Apriori 6.5. Khám phá các kết hợp dựa trên ràng buộc 6.6. Phân tích tương quan 6.7. Tóm tắt	[1-6,8]	Hiểu, Năm vững, Vận dụng, Tổng hợp
11, 12	Chương 7: Khai phá dữ liệu và công nghệ cơ sở 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	[1,2,5-8]	Vận dụng, Tổng hợp
12, 13	Chương 8: Ứng dụng khai phá dữ liệu 8.1. Tổng quan về khả năng ứng dụng của khai phá dữ liệu 8.2. Ứng dụng khai phá dữ liệu trong lĩnh vực y học 8.3. Ứng dụng khai phá dữ liệu trong lĩnh vực khoa học 8.4. Ứng dụng khai phá dữ liệu trong lĩnh vực kinh tế 8.5. Ứng dụng khai phá dữ liệu trong lĩnh vực tài chính 8.6. Ứng dụng khai phá dữ liệu trong lĩnh vực thương mại 8.7. Tóm tắt	[1,4,8]	Hiểu, Năm vững, Vận dụng, Tổng hợp
14	Chương 9: Các đề tài nghiên cứu trong khai phá dữ liệu 9.1. Hướng dữ liệu 9.2. Hướng kỹ thuật 9.3. Hướng ứng dụng 9.4. Tóm tắt	[1-8]	Hiểu, Năm vững, Vận dụng, Tổng hợp
15	Chương 10: Ôn tập 10.1. Quá trình khai phá dữ liệu 10.2. Các vấn đề chính trong khai phá dữ liệu 10.3. Tóm tắt	[1-8]	Hiểu, Năm vững, Vận dụng, Tổng hợp

- **PHẦN BÀI TẬP, TIỂU LUẬN NGOẠI KHÓA, HV ĐI NGHIÊN CỨU THỰC TIỄN NGOÀI TRƯỜNG:** 15 tiết

TT	Nội dung	Số tiết	Địa điểm	TLTK
1	Các giải thuật khai phá dữ liệu không gian	3		

2	Các giải thuật khai phá dữ liệu thời gian	3		
3	Các giải thuật khai phá dữ liệu đa phương tiện	3		
4	Các giải thuật khai phá dữ liệu web	3		
5	Các giải thuật khai phá dữ liệu văn bản	3		

Tp.Hồ Chí Minh, ngày tháng năm 2013

**KT.TRƯỞNG KHOA
PHÓ TRƯỞNG KHOA**

CB PHỤ TRÁCH LẬP ĐỀ CƯƠNG

TS. PHẠM TRẦN VŨ

TS. VÕ THỊ NGỌC CHÂU