

Ch ̣ng tr ̣nh Giáo d ̣c i h c
Ngành ào t ờ: H th ng thông tin Tr ̣nh ào t ờ: i h c
Ch ̣ng tr ̣nh ào t ờ: Công ngh ̣ th ng tin

c ̣ng chi ti t h c ph n

1. Tên h c ph n: Khai phá d ̣ li u

Mã h c ph n: DAMI330484

2. Tên Ti ng Anh: Datamining

3. S ̣ t ̣n ch : 3(2+1)

Phân b ̣ th i gian: (h c k 15 tu n) 3(2:1:6)

5. Các gi ng viên ph ̣ trách h c ph n:

1/ GV ph ̣ trách chính: ThS. Nguy n Thanh Tu n

2/ Danh sách gi ng viên cùng GD:

6. i u ki n tham gia h c t p h c ph n

Môn h c tiên quy t:

- + C s l p tr ̣nh
- + C u trúc d ̣ li u và gi i thu t
- + Xác su t th ng kê
- + C s d ̣ li u
- + H qu n tr CSDL

7. Mô t tóm t t h c ph n

H c ph n này trang b ̣ cho ng i h c nh ng ki n th c c b n v khai phá d ̣ li u; các khái ni m có liên quan, ý ngh a và t m quan tr ̣ng. H c ph n cung c p cho ng i h c ki n th c và k n ng trong phân lo i, ánh giá các h th ng khai phá d ̣ li u. H c ph n trang b ̣ các k thu t tính toán m i và thông d ̣ng nh t phân lo i, trích l c, ánh giá các thông tin trong quá tr ̣nh h ̣ tr ra quy t ̣nh cho các t ̣ch c kinh t , giáo d ̣c, y t ... Ngoài ra, h c ph n còn h ̣ng d ̣n ng i h c s d ̣ng các công c mã ngu n m , các kho d ̣ li u h ̣ tr trong quá tr ̣nh cài t và th nghi m h th ng khai phá d ̣ li u.

8. M c tiêu h c ph n (Course objective)

M c tiêu (Goals)	Mô t (Goal description) <i>(H c ph n này trang b ̣ cho sinh viên:)</i>	Chu n u ra CT T
G1	Ki n th c khai phá d ̣ li u	1.2, 1.3
G2	Hi u và tr ̣nh bày các k thu t tính toán và các thu t toán khai phá d ̣ li u	2.1, 2.2
G3	K n ng làm vi c nh ̣m, và thuy t tr ̣nh	3.1, 3.2

G4	Kh n ng v n d ng ki n th c áp d ng các thu t toán khai phá d li u th c t	4.1, 4.3
-----------	--	----------

9. Chu n u ra c a h c ph n

M c tiêu	Chu n u ra h c ph n	Mô t <i>(Sau khi h c xong môn h c này, ng i h c có th :)</i>	Chu n u ra CDIO
G1	G1.1	Trình bày các khái ni m c b n và gi i thích c các thu t ng c a khai phá d li u	1.2
	G1.2	Hi u c ý ngh a, t m quan tr ng và các thách th c c a khai phá d li u	1.2
	G1.3	Trình bày c các c tính c a d li u nh : m c l p l i, liên k t, phát sinh lu t k t h p	1.3
	G1.4	Hi u c quy trình và k thu t khai phá d li u	1.3
	G1.5	Trình bày c các thu t toán khai phá d li u	1.3
	G1.6	Hi u bi t và tính toán các o trong các ph ng pháp ánh giá k t qu phân l p và gom c m d li u	1.3
	G1.7	Hi u bi t và phân lo i các h th ng khai phá d li u và xu h ng phát tri n	1.3
G2	G2.1	S d ng s d ng c các tools mã ngu n m , thu t toán phân l p và gom c m d li u	2.1
	G2.2	Áp d ng các k thu t, thu t toán phân l p và gom c m d li u th nghi m trên các kho d li u	2.2
G3	G3.1	Nâng cao k n ng phân công công vi c và làm vi c theo nhóm	3.1
	G3.2	Trình bày tr c ám ông s d ng ph ng tí n trình chi u	3.2
G4	G4.1	ánh giá và l a ch n các k thu t thích h p v i d li u th c t	4.1, 4.4

10. Nhi m v c a sinh viên

SV không th c hi n m t trong các nhi m v sau ây s b c m thi:

- D l p: t i thi u 80% s tí t gi ng
- Bài t p: ph i hoàn thành 100% bài t p v nhà
- Bài thí nghi m: ph i hoàn thành 100% thí nghi m mô ph ng các k thu t

11. Tài li u h c t p

- Sách, giáo trình chính:
 - + Data Mining: Concepts and Techniques (2nd), Jiawei Han, Diane Cerra, 2006
- Sách (TLTK) tham kh o:
 - + Top 10 algorithms in data mining, Xindong Wu et al, Knowl Inf Syst, 2008

12. T l Ph n tr m các thành ph n i m và các hình th c ánh giá sinh viên:

- Thang i m: 10
- K ho ch ki m tra nh sau:

Hình thức KT	Nội dung	Thời gian	Công cụ KT	Chuẩn ra KT	Tỉ lệ (%)
Bài tập					50
BT#1	Thí nghiệm tools mã nguồn mở	Tuần 4	Bài tập nh trên lớp	G1.1, G1.4, G2.3	10
BT#2	Bài tập thuật toán phân lớp dữ liệu	Tuần 6	Bài tập nh trên lớp	G2.2, G2.4	20
BT#3	Bài tập thuật toán gom cụm dữ liệu	Tuần 7	Bài tập nh trên lớp	G2.2, G2.4	20
Thi cuối kỳ					50
	- Nội dung bao quát tất cả các chuẩn ra quan trọng của môn học. - Thời gian làm bài 75-90 phút.		Thi tự luận	G2.2	

13. Nội dung chi tiết học phần

Tuần	Nội dung	Chuẩn ra học phần
1	Chương 1: Tổng quan và khai phá dữ liệu	
	A/ Các nội dung và PPGD chính trên lớp: (4)	G1.1, G1.4, G2.3
	Nội Dung GD chính trên lớp: + Các khái niệm cơ bản + Phân loại các hệ thống khai phá dữ liệu + Metrics và thách thức PPGD chính: + Thuyết giảng + Minh họa các hệ thống khai phá dữ liệu + Trình chiếu Powerpoint	
	B/ Các nội dung content nhà: (8)	
	+ content thêm: Chức năng và nhiệm vụ của datamining + Làm các bài tập case giao + Tài liệu: Data Mining: Concepts and Techniques, chương 1.	
2	Chương 2: Tỉ lệ xử lý dữ liệu	
	A/ Tóm tắt các ND và PPGD chính trên lớp: (4) Nội Dung GD chính trên lớp: + Tóm quan trọng của tỉ lệ xử lý dữ liệu	G1.1, G1.4, G2.3

	<ul style="list-style-type: none"> + Kỹ thuật làm sạch dữ liệu (data cleaning) + Kỹ thuật tích hợp và chuyển đổi dữ liệu + Kỹ thuật thu gom kích thước dữ liệu <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa các loại dữ liệu và kỹ thuật xử lý dữ liệu + Trình chiếu Powerpoint 	
	<p>B/ Các nội dung cốt lõi của nhà: (8)</p> <ul style="list-style-type: none"> + Xem lại và tóm tắt ngắn gọn các kỹ thuật xử lý + Làm các bài tập case giao + Tài liệu: Data Mining: Concepts and Techniques, chương 2 	
3	<p><i>Thực hành chương 2</i></p> <p>Thí nghiệm tools mã nguồn mở</p>	
	<p>A/ Các nội dung và PPGD chính trên lớp: (4)</p> <p>Nội dung GD:</p> <ul style="list-style-type: none"> + Hướng dẫn thí nghiệm tools mã nguồn mở <p>PPGD chính:</p> <ul style="list-style-type: none"> + Làm mẫu. + Tác động hỗ trợ sinh viên 	G1.1, G1.4, G3.1
	<p>B/ Các nội dung cốt lõi của nhà: (8)</p> <ul style="list-style-type: none"> + Công cụ bài tập đã làm trên lớp 	
4	<p>Chương 3: Khai phá các mẫu, liên kết và quan hệ trong dữ liệu</p>	
	<p>A/ Tóm tắt các ND và PPGD chính trên lớp: (4)</p> <p>Nội Dung GD chính trên lớp:</p> <ul style="list-style-type: none"> + Các khái niệm cơ bản + Khai phá các mẫu dữ liệu thông xuyên + Khai phá các luật kết hợp + Phân tích tương quan (correlation analysis) <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Trình chiếu Powerpoint 	G1.1, G1.4,
	<p>B/ Các nội dung cốt lõi của nhà: (8)</p> <ul style="list-style-type: none"> + Xem lại và tóm tắt ngắn gọn nội dung của chương + Làm các bài tập case giao + Tài liệu: Data Mining: Concepts and Techniques, chương 5 	
5, 6	<p>Chương 4: Phân loại dữ liệu</p>	
	<p>A/ Tóm tắt các ND và PPGD chính trên lớp: (8)</p> <p>Nội Dung GD chính trên lớp:</p> <ul style="list-style-type: none"> + Các khái niệm cơ bản + Các công cụ mã nguồn mở (SVM linear, SVM light, Weka) 	G2.2, G2.4, G3.1

	<ul style="list-style-type: none"> + Các loại mô hình và phương pháp đánh giá + Kỹ thuật sử dụng cây quyết định + Kỹ thuật Bayesian + Kỹ thuật sử dụng luật (rule-based) + Kỹ thuật lai tạo + Kỹ thuật SVM + Kỹ thuật phân tích luỹ kết hợp + Kỹ thuật lazy learner (học theo lần học) <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm 	
	<p>B/ Các nội dung cần thực hiện: (16)</p> <ul style="list-style-type: none"> + Xem thêm các ví dụ về lý của các kỹ thuật + Download các công cụ (tools) nguồn mở + Download các kho dữ liệu thử nghiệm + Các tài liệu hướng dẫn sử dụng các tools và kho dữ liệu + Làm các bài tập thực hành + Thử nghiệm dữ liệu trên các tools + Báo cáo kết quả (nộp file trên mạng cho giảng viên) <p>Tài liệu:</p> <ul style="list-style-type: none"> + Data Mining: Concepts and Techniques, chương 6 + Các tài liệu hướng dẫn sử dụng tools và kho dữ liệu 	
7, 8	<p><i>Thực hành chương 3, 4:</i></p> <p>Bài tập thuật toán phân loại dữ liệu</p>	
	<p>A/ Các nội dung và PPGD chính trên lớp: (4)</p> <p>Nội dung GD:</p> <ul style="list-style-type: none"> + Hướng dẫn bài tập thuật toán phân loại dữ liệu <p>PPGD chính:</p> <ul style="list-style-type: none"> + Làm mẫu. + Tác động hỗ trợ và sinh viên 	G2.2, G2.4
	<p>B/ Các nội dung cần thực hiện: (8)</p> <ul style="list-style-type: none"> + Công việc bài tập đã làm trên lớp 	
9, 10	<p>Chương 5: Gom cụm dữ liệu</p> <p>A/ Tóm tắt các ND và PPGD chính trên lớp: (8)</p> <p>Nội Dung GD chính trên lớp:</p> <ul style="list-style-type: none"> + Các khái niệm cơ bản + Kỹ thuật phân chia (Partitioning) + Kỹ thuật phân cấp (Hierarchical) + Kỹ thuật density-based 	G2.2, G2.4

	<ul style="list-style-type: none"> + Kỹ thuật số lưới (Grid-based) + Kỹ thuật số mô hình (Model-based) <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm 	
	<p>B/ Các nội dung cần thực hiện: (16)</p> <ul style="list-style-type: none"> + Xem thêm các ví dụ về lý của các kỹ thuật + Download các công cụ (tools) nguồn + Download các kho dữ liệu thử nghiệm + Các tài liệu hướng dẫn sử dụng các tools và kho dữ liệu + Làm các bài tập thực giao + Thử nghiệm dữ liệu trên các tools + Báo cáo kết quả (nộp file trên mạng cho giảng viên) + Tài liệu: + Data Mining: Concepts and Techniques, chương 7 + Các tài liệu hướng dẫn sử dụng tools và kho dữ liệu 	
11, 12	<p><i>Thực hành chương 3, 4:</i></p> <p>Bài tập thu thập toán học mô dữ liệu</p>	
	<p>A/ Các nội dung và PPGD chính trên lớp: (4)</p> <p>Nội dung GD:</p> <ul style="list-style-type: none"> + Hướng dẫn bài tập thu thập toán phân lập dữ liệu <p>PPGD chính:</p> <ul style="list-style-type: none"> + Làm mẫu. + Tác động hỗ trợ và sinh viên 	G2.2, G2.4, G3.1
	<p>B/ Các nội dung cần thực hiện: (8)</p> <ul style="list-style-type: none"> + Công việc bài tập đã làm trên lớp 	
13	<p>Chương 6: Ứng dụng và xu hướng phát triển</p> <p>A/ Tóm tắt các ND và PPGD chính trên lớp: (4)</p> <p>Nội Dung GD chính trên lớp:</p> <ul style="list-style-type: none"> + Các ứng dụng khai phá dữ liệu + Phát triển và hoàn thiện luật + Khai phá dữ liệu tác động xã hội + Xu hướng phát triển <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa các ứng dụng + Trình chiếu Powerpoint + Thảo luận nhóm 	G1.1, G1.4, G2.3

	<p>B/ Các nội dung cần thực hiện: (8)</p> <ul style="list-style-type: none"> + Xem thêm: Khai phá các dữ liệu thời gian, time-series, data streams + Làm các bài tập giao + Tài liệu: Data Mining: Concepts and Techniques, chương 11. 	
14, 15	<p>Chương 7: Thuật toán phân loại và ứng dụng</p> <p>A/ Tóm tắt các ND và PPGD chính trên lớp: (4)</p> <p>Nội Dung (ND) chính trên lớp:</p> <ul style="list-style-type: none"> + Thuật toán C4.5 and beyond + Thuật toán k-means + Thuật toán SVM (Support vector machines) <p>PPGD chính:</p> <ul style="list-style-type: none"> + Thuyết giảng + Minh họa các thuật toán + Trình chiếu Powerpoint + Thảo luận nhóm 	G1.1, G1.4, G2.3, G2.2, G2.4
	<p>B/ Các nội dung cần thực hiện: (8)</p> <ul style="list-style-type: none"> + Xem thêm các ví dụ về lý các thuật toán + Download các công cụ (tools) nghiên cứu + Download các kho dữ liệu nghiên cứu + Xem tài liệu hướng dẫn sử dụng các tools và kho dữ liệu + Làm các bài tập giao + Thực hiện dự án trên các tools + Báo cáo kết quả (nộp file trên mạng cho giảng viên) + Tài liệu: + Các tài liệu hướng dẫn sử dụng tools và kho dữ liệu + Top 10 algorithms in data mining 	

14. Yêu cầu khoa học:

- + Các bài tập và bài thực nghiệm phải thực hiện chính bản thân sinh viên. Nếu phát hiện có sao chép thì xử lý các sinh viên có liên quan bằng hình thức đánh giá 0 (không) điểm và cấm thi cử tiếp.
- + Không có thi học kỳ. Nếu phát hiện thì sẽ bị kỷ luật theo quy định của Khoa và nhà trường.

15. Ngày phê duyệt lần cuối:

16. Các thành viên:

Trưởng khoa

Trưởng BM

Nhóm biên soạn

TS. Nguyễn Thanh Dũng

TS. Nguyễn Thành Sơn

ThS. Nguyễn Thanh Tuấn

17. Tiến trình cập nhật CCT

Lần 1: Nội Dung Cập nhật CCT lần 1: ngày tháng năm	<nguyên tắc cập nhật kỹ
---	-------------------------

	<p>và ghi rõ họ tên)</p> <p>Tờ trình Bộ môn:</p>
<p>L n 2: Nội Dung Công trình CCT l n 2: ngày tháng năm</p>	<p><ng i c p nh t ký và ghi rõ họ tên)</p> <p>Tờ trình Bộ môn:</p>