

TRƯỜNG ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH  
KHOA CÔNG NGHỆ THÔNG TIN

**ĐỀ CƯƠNG MÔN HỌC**

**I. Thông tin tổng quát**

1. Tên môn học tiếng Việt: KHAI PHÁ DỮ LIỆU
2. Tên môn học tiếng Anh: DATA MINING
3. Thuộc khối kiến thức/kỹ năng
  - Giáo dục đại cương
  - Kiến thức chuyên ngành
  - Kiến thức cơ sở
  - Kiến thức bổ trợ
  - Kiến thức ngành
  - Đồ án/Khóa luận tốt nghiệp
4. Số tín chỉ

Tổng số	Lý thuyết	Thực hành	Tự học
3	2	1	3(2, 1, 5)

5. Phụ trách môn học
  - a) Khoa/Ban/Bộ môn: Công nghệ Thông tin
  - b) Giảng viên: TS. Trương Hoàng Vinh
  - c) Địa chỉ email liên hệ: vinh.th@ou.edu.vn
  - d) Phòng làm việc: 604

**II. Thông tin về môn học**

1. Mô tả môn học

Khai phá dữ liệu là quá trình khám phá các tri thức mới và các tri thức có ích ở dạng tiềm năng trong nguồn dữ liệu đã có. Mục tiêu của khai phá dữ liệu là sử dụng các giải thuật xử lý dữ liệu để biến dữ liệu thô thành dữ liệu có cấu trúc dễ hiểu để sử dụng tiếp. Các giải thuật này được tổng hợp từ nhiều lĩnh vực thuộc khoa học máy tính như trí tuệ nhân tạo, máy học, thống kê và hệ thống cơ sở dữ liệu.

2. Môn học điều kiện

STT	Môn học điều kiện	Mã môn học
1.	Môn tiên quyết	
	Không	
2.	Môn học trước	

STT	Môn học điều kiện	Mã môn học
	Cơ sở dữ liệu	ITEC2502
	Trí tuệ nhân tạo	ITEC3413
3.	Môn học song hành	
	Không	

### 3. Mục tiêu môn học

Sinh viên học xong môn học có khả năng:

Mục tiêu môn học	Mô tả	CĐR CTĐT phân bổ cho môn học
CO1	<ul style="list-style-type: none"> <li>- Hiểu được khái niệm của khai phá dữ liệu, các thuận lợi và giới hạn của nó.</li> <li>- Hiểu được các thuật toán khai phá dữ liệu phổ biến.</li> <li>- Hiểu được cách khai thác tri thức, thông tin từ dữ liệu.</li> <li>- Hiểu khai phá dữ liệu trong khoa học máy tính cũng như ứng dụng thực tế.</li> <li>- Hiểu được các hướng nghiên cứu và ứng dụng hiện nay về khai phá dữ liệu.</li> </ul>	PLO6.14
CO2	<ul style="list-style-type: none"> <li>- Vận dụng và cài đặt được các thuật toán khai phá dữ liệu phổ biến.</li> <li>- Vận dụng kiến thức khai phá dữ liệu và phát triển các ứng dụng thực tế.</li> <li>- Có khả năng nghiên cứu thêm lý thuyết các thuật toán khai phá dữ liệu.</li> </ul>	PLO3.1 PLO4.1 PLO4.2 PLO6.14
CO3	<ul style="list-style-type: none"> <li>- Tinh thần tự học, tự nghiên cứu.</li> <li>- Không ngừng cập nhật những kết quả nghiên cứu mới lĩnh vực khai phá dữ liệu.</li> </ul>	PLO12.3

### 4. Chuẩn đầu ra (CĐR) môn học

Học xong môn học này, sinh viên làm được (đạt được):

Mục tiêu môn học	CĐR môn học	Mô tả CĐR
CO1	CLO1.1	Hiểu được tổng quan khai phá dữ liệu, các thuật toán khai phá dữ liệu phổ biến.
	CLO1.2	Hiểu được tầm quan trọng của khai phá dữ liệu.
	CLO1.3	Hiểu được các hướng nghiên cứu và ứng dụng hiện nay về khai phá dữ liệu.
CO2	CLO2.1	Cài đặt được các thuật toán khai phá dữ liệu.
	CLO2.2	Vận dụng các kiến thức khai phá dữ liệu phát triển các ứng dụng thực tế.

Mục tiêu môn học	CĐR môn học	Mô tả CĐR
	CLO2.3	Có khả năng nghiên cứu thêm lý thuyết các thuật toán khai phá dữ liệu.
CO3	CLO3.1	Nâng cao khả năng tự học, tự nghiên cứu.
	CLO3.2	Thúc đẩy cập nhật kiến thức, kết quả nghiên cứu mới trong khai phá dữ liệu.

Ma trận tích hợp giữa chuẩn đầu ra của môn học và chuẩn đầu ra của chương trình đào tạo

CLOs	PLO3.1	PLO4.1	PLO4.2	PLO6.14	PLO12.3
<b>1.1</b>				4	
<b>1.2</b>				4	
<b>1.3</b>				4	
<b>2.1</b>	3	3	3	5	
<b>2.2</b>	3	3	3	5	
<b>2.3</b>				5	
<b>3.1</b>					5
<b>3.2</b>					5

1: Không đáp ứng

2: Ít đáp ứng

3: Đáp ứng trung bình

4: Đáp ứng nhiều

5: Đáp ứng rất nhiều

## 5. Học liệu

### a) Giáo trình

[1] Charu C. Aggarwal. Data Mining: The Textbook. Springer. 2015. [49465]

[2] Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis Fundamental Concepts and Algorithms. Cambridge University Press. 2014. [49473]

### b) Tài liệu tham khảo (liệt kê tối đa 3 tài liệu tham khảo)

[3] Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman. Mining of Massive Datasets (2nd edition). Cambridge University Press. 2014. [49466]

### c) Phần mềm

1) Python 3.7.4

2) Pycharm Community

## 6. Đánh giá môn học

Thành phần đánh giá	Bài đánh giá	Thời điểm	CDR môn học	Tỷ lệ %
(1)	(2)	(3)	(4)	
A1. Đánh giá quá trình	A.1.1. Đánh giá trên lớp/chuyên cần		CLO1.2, CLO1.2, CLO1.3	10%
	Tổng cộng: 01			10%
A2. Đánh giá giữa kỳ	A2.1. Bài tập lớn		CLO1.2, CLO1.2, CLO1.3	30%
	Tổng cộng: 01			30%
A3. Đánh giá cuối kỳ	A3.1. Thi viết trên giấy		CLO1.2, CLO1.2, CLO1.3, CLO2.2, CLO2.3	60%
	Tổng cộng: 01			60%
Tổng cộng				<b>100%</b>

## 7. Kế hoạch giảng dạy

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
1. Tuần 1/ Lý thuyết 1	Chương 1. Tổng quan về khai phá dữ liệu 1.1. Tổng quan khai phá dữ liệu 1.2. Dạng dữ liệu và mẫu cần tập trung khai phá 1.3. Các phương pháp và ứng dụng cho bài toán khai phá dữ liệu 1.4. Các yếu tố ảnh hưởng	PO1.1 PO1.2 PO1.3	<p>Giảng viên: + Giới thiệu đề cương chi tiết. + Thuyết giảng + Đặt câu hỏi, bài tập. + Nhấn mạnh những điểm chính. + Nêu các yêu cầu cho buổi học sau.</p> <p>Sinh viên: + Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú. + Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. + Trên hệ thống LMS: trả lời các câu</p>	A1.1 A2.1 A3.1	[1]

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
			hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.		
1) Tuần 2/ Lý thuyết 2	<p>Chương 2. Tiền xử lý dữ liệu</p> <p>2.1. Khảo sát dữ liệu</p> <p>2.2. Thống kê</p> <p>2.3. Hình tượng hóa</p> <p>2.4. Đo đặc độ tương đồng</p> <p>2.5. Làm sạch dữ liệu</p> <p>2.5.1. Xử lý dữ liệu bị thiếu</p> <p>2.5.2. Xử lý dữ liệu không đúng và không nhất quán</p> <p>2.5.3. Chuẩn hóa và mở rộng quy mô</p> <p>2.6. Tích hợp và thu gọn dữ liệu</p> <p>2.6.1. Chọn mẫu</p> <p>2.6.2. Lựa chọn tập con đặc trưng</p> <p>2.6.3. Giảm chiều với trục xoay</p> <p>2.6.4. Giảm chiều với chuyển đổi kiểu</p> <p>2.7. Biến đổi dữ liệu và rời rạc hóa</p>	PO2.1 PO2.2 PO2.3	<p>Giảng viên:</p> <ul style="list-style-type: none"> <li>+ Giới thiệu đề cương chi tiết.</li> <li>+ Thuyết giảng</li> <li>+ Đặt câu hỏi, bài tập.</li> <li>+ Nhấn mạnh những điểm chính.</li> <li>+ Nêu các yêu cầu cho buổi học sau.</li> </ul> <p>Sinh viên:</p> <ul style="list-style-type: none"> <li>+ Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú.</li> <li>+ Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan.</li> <li>+ Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.</li> </ul>	A1.1 A2.1 A3.1	[1]
2) Tuần 3/ Lý thuyết 3	<p>Chương 3. Khai thác các mẫu phổ biến</p> <p>3.1. Khai thác tập hợp</p> <p>3.1.1. Tập phổ biến và luật kết hợp</p> <p>3.1.2. Các giải thuật khai thác tập hợp</p> <p>3.1.3. Các luật kết hợp tổng quát</p> <p>3.2. Khai thác sự tuần tự</p> <p>3.2.1. Trình tự phổ biến</p> <p>3.2.2. Khai thác các trình tự phổ biến</p> <p>3.2.3. Khai thác các chuỗi con thông qua các cây hậu tố</p>	PO2.1 PO2.2 PO2.3	<p>Giảng viên:</p> <ul style="list-style-type: none"> <li>+ Giới thiệu đề cương chi tiết.</li> <li>+ Thuyết giảng</li> <li>+ Đặt câu hỏi, bài tập.</li> <li>+ Nhấn mạnh những điểm chính.</li> <li>+ Nêu các yêu cầu cho buổi học sau.</li> </ul> <p>Sinh viên:</p> <ul style="list-style-type: none"> <li>+ Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú.</li> </ul>	A1.1 A2.1 A3.1	[1][3]

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
	3.3. Khai thác các mẫu Graph 3.3.1. Hỗ trợ và đăng cấu 3.3.2. Khởi tạo các ứng viên 3.3.3. Thuật toán gSpan		+ Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. +Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.		
3) Tuần 4/ Lý thuyết 4	Chương 4. Phân loại dữ liệu 4.1. Nền tảng toán học 4.2. Cây quyết định 4.2.1. Các cây quyết định 4.2.2. Các tiêu chí để tách 4.2.3. Cắt tia cây và tiêu chí dừng 4.3. Thuật toán Bayes	PO2.1 PO2.2 PO2.3 PO3.1 PO3.2	Giảng viên: + Giới thiệu đề cương chi tiết. + Thuyết giảng + Đặt câu hỏi, bài tập. + Nhấn mạnh những điểm chính. + Nêu các yêu cầu cho buổi học sau.  Sinh viên: + Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú. + Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. +Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.	A3.1	[1][2]
4) Tuần 5/ Lý thuyết 5	Chương 4. Phân loại dữ liệu (tt) 4.4. Phân loại dữ liệu dựa vào các luật 4.4.1. Khởi tạo các luật từ các cây quyết định 4.4.2. Các giải thuật bao phủ tuần tự 4.4.3. Cắt tia các luật	PO2.1 PO2.2 PO2.3 PO3.1 PO3.2	Giảng viên: + Giới thiệu đề cương chi tiết. + Thuyết giảng + Đặt câu hỏi, bài tập. + Nhấn mạnh những điểm chính. + Nêu các yêu cầu	A3.1	[1][2]

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
	4.4.4. Các phân loại kết hợp 4.5. Các kỹ thuật nâng cao độ chính xác.		cho buổi học sau.  Sinh viên: + Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú. + Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. +Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.		
5) Tuần 6/ Lý thuyết 6	Chương 5. Gồm cụm dữ liệu 5.1. Nền tảng toán học 5.2. Phân hoạch k-mean 5.3. Gồm cụm dựa trên cấu trúc lồng nhau	PO2.1 PO2.2 PO2.3 PO3.1 PO3.2	Giảng viên: + Giới thiệu đề cương chi tiết. + Thuyết giảng + Đặt câu hỏi, bài tập. + Nhấn mạnh những điểm chính. + Nêu các yêu cầu cho buổi học sau.  Sinh viên: + Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú. + Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. +Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.	A3.1	[1][2]
6) Tuần 7/	Chương 5. Gồm cụm dữ liệu (tt)	PO2.1 PO2.2	Giảng viên: + Giới thiệu đề	A3.1	[1][2]

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
Lý thuyết 7	5.4. Gom cụm dựa trên lưới 5.5. Đánh giá độ hiệu quả của việc gom cụm	PO2.3 PO3.1 PO3.2	<p>ương chi tiết. + Thuyết giảng + Đặt câu hỏi, bài tập. + Nhấn mạnh những điểm chính. + Nêu các yêu cầu cho buổi học sau.</p> <p>Sinh viên: + Học ở lớp: nghe giảng, trả lời các câu hỏi, giải các bài tập đặt ra, ghi chú. + Học ở nhà: xem bài giảng, đúc kết các kiến thức trọng tâm, tìm hiểu các kiến thức liên quan. + Trên hệ thống LMS: trả lời các câu hỏi trắc nghiệm lý thuyết, tham gia thảo luận trên diễn đàn.</p>		
7) Tuần 1/ Thực hành 1	Các kỹ thuật thu thập dữ liệu	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên: Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên: + Làm lại các bài mẫu giảng viên demo/hướng dẫn. + Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</p>	A2.1 A2.2	
8) Tuần 2/ Thực hành 2	Các kỹ thuật tiền xử lý dữ liệu	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên: Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm</p>	A2.1 A2.2	



Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
			<p>các bài cần hướng dẫn.</p> <p>Sinh viên:</p> <ul style="list-style-type: none"> <li>+ Làm lại các bài mẫu giảng viên demo/hướng dẫn.</li> <li>+ Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</li> </ul>		
9) Tuần 3/ Thực hành 3	Các thuật toán phân loại dữ liệu	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên:</p> <p>Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên:</p> <ul style="list-style-type: none"> <li>+ Làm lại các bài mẫu giảng viên demo/hướng dẫn.</li> <li>+ Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</li> </ul>	A2.1 A2.2	
10) Tuần 4/ Thực hành 4	Các thuật toán phân loại dữ liệu (tt)	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên:</p> <p>Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên:</p> <ul style="list-style-type: none"> <li>+ Làm lại các bài mẫu giảng viên demo/hướng dẫn.</li> <li>+ Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</li> </ul>	A2.1 A2.2	

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
11) Tuần 5/ Thực hành 5	Các thuật toán gom cụm dữ liệu	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên: Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên: + Làm lại các bài mẫu giảng viên demo/hướng dẫn. + Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</p>	A2.1 A2.2	
12) Tuần 6/ Thực hành 6	Các thuật toán gom cụm dữ liệu (tt)	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên: Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên: + Làm lại các bài mẫu giảng viên demo/hướng dẫn. + Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.</p>	A2.1 A2.2	
13) Tuần 7/ Thực hành 7	Các phương pháp thực nghiệm	PO2.1 PO2.2 PO3.1 PO3.2	<p>Giảng viên: Demo chương trình mẫu, hoặc hướng dẫn sơ qua cách làm các bài cần hướng dẫn.</p> <p>Sinh viên: + Làm lại các bài mẫu giảng viên demo/hướng dẫn.</p>	A2.1 A2.2	

Tuần/buổi học	Nội dung	CDR môn học	Hoạt động dạy và học	Bài đánh giá	Tài liệu chính và tài liệu tham khảo
(1)	(2)	(3)	(4)	(5)	(6)
			+ Tự làm các bài tập tự làm dựa trên các demo mẫu và lý thuyết đã học.		

8. Quy định của môn học

- Sinh viên tham gia đầy đủ các buổi học lý thuyết và thực hành.
- Sinh viên phải nộp bài tập lớn thông qua hệ thống LMS và tham gia vấn đáp.

**TRƯỞNG KHOA**

*(Đã ký)*

TS.GVCC. Lê Xuân Trường

**GIẢNG VIÊN BIÊN SOẠN**

*(Đã ký)*

TS. Trương Hoàng Vinh