

TP. Hồ Chí Minh, ngày tháng năm

ĐỀ CƯƠNG MÔN HỌC

PHÂN TÍCH DỮ LIỆU CHO TÀI CHÍNH

A. THÔNG TIN CHUNG VỀ MÔN HỌC

- 1. Tên môn học (tiếng Việt) : PHÂN TÍCH DỮ LIỆU CHO TÀI CHÍNH**
- 2. Tên môn học (tiếng Anh) : DATA ANALYTICS IN FINANCE**
- 3. Mã số môn học : DAT709**
- 4. Trình độ đào tạo : Đại học**
- 5. Ngành đào tạo áp dụng : Tất cả các ngành đào tạo đại học**
- 6. Số tín chỉ : 03**
 - Lý thuyết : 02
 - Thảo luận và bài tập : 00
 - Thực hành : 01
 - Khác (ghi cụ thể) : Tự học, bài tập cá nhân, bài tập nhóm
- 7. Phân bổ thời gian :**
 - Tại giảng đường : 60 tiết
 - Tự học ở nhà : 90 giờ
 - Trực tuyến : Giảng viên có thể bố trí học online nhưng tổng số không quá 30% số tiết của toàn môn học.
 - Khác (ghi cụ thể) : Đọc tài liệu, làm bài về nhà, làm bài tập cá nhân và bài tập nhóm chiếm tối thiểu 2 lần so với thời gian học tập trên lớp.
- 8. Khoa quản lý môn học : Bộ môn Toán Kinh tế**
- 9. Môn học trước : Không**
- 10. Mô tả môn học :**

Môn học thuộc khối kiến thức chuyên ngành và là môn học bắt buộc trong chương trình đào tạo. Môn học cung cấp cách thức phân tích dữ liệu. Ngày nay, dữ liệu được thu thập một cách liên tục, rộng khắp dần dẫn đến sự gia tăng nhanh chóng về khối lượng dữ liệu cũng như sự đa dạng và hỗn tạp về định dạng. Chiết xuất thông tin có giá trị từ một khối lượng dữ liệu lớn, hỗn tạp như vậy là thách thức lớn không

nhỏ cho các giải pháp phân tích dữ liệu cổ điển. Những điều này đòi hỏi phải có cách tiếp cận dữ liệu một cách hiệu quả, thông minh, nhanh chóng và chính xác hơn. Môn học khai phá dữ liệu cung cấp kiến thức nền tảng về các thuật ngữ nhập môn lĩnh vực khai thác dữ liệu; các vấn đề liên quan đến việc tiền xử lý dữ liệu; phương pháp phân loại dữ liệu, gom cụm dữ liệu và khai thác mẫu; ứng dụng của khai thác dữ liệu trong trường hợp cụ thể, với dạng dữ liệu có cấu trúc phức tạp như dòng dữ liệu, dữ liệu chuỗi thời gian, dữ liệu mạng xã hội.

11. Mục tiêu và chuẩn đầu ra của môn học

11.1. Mục tiêu của môn học

Mục tiêu	Mô tả mục tiêu	Nội dung CĐR CTĐT ¹ phân bổ cho môn học	CĐR CTĐT
(a)	(b)	(c)	(d)
CO1	Nhận ra ý nghĩa của khai phá dữ liệu. Trình bày được khái niệm liên quan đến khai phá dữ liệu, các kỹ thuật xây dựng khai phá dữ liệu.	Kỹ năng thu thập, phân tích và xử lý thông tin	PLO5
CO2	Hiểu được các quy trình và các phương pháp khai phá dữ liệu và có thể áp dụng và giải quyết các bài toán trong thực tiễn.	Khả năng ứng dụng nghiên cứu trong thực tiễn. Kỹ năng thu thập, phân tích và xử lý thông tin . Kỹ năng sử dụng máy tính	PLO1
CO3	Phát hiện được các vấn đề trong Khai phá dữ liệu. Phân tích và xử lý dữ liệu trong quá trình khai phá dữ liệu. Sử dụng tư duy sáng tạo và tư duy phản biện ở mức độ đơn giản cho vấn đề ứng dụng khai phá dữ liệu.	Khả năng ứng dụng nghiên cứu trong thực tiễn. Kỹ năng thu thập, phân tích và xử lý thông tin . Kỹ năng sử dụng máy tính	PLO1 PLO2

11.2. Chuẩn đầu ra của môn học (CĐR MH) và sự đóng góp vào chuẩn đầu ra của chương trình đào tạo (CĐR CTĐT)

¹ Giải thích ký hiệu viết tắt: CĐR – chuẩn đầu ra; CTĐT - chương trình đào tạo.

CĐR MH	Nội dung CĐR MH	Mức độ theo thang đo của CĐR MH	Mục tiêu môn học	CĐR CTĐT
(a)	(b)	(c)	(d)	(e)
CLO1	Định nghĩa thuật ngữ liên quan đến khai phá dữ liệu. Phát biểu được các đặc trưng của khai phá dữ liệu. Trình bày được quá trình phát triển của khai phá dữ liệu	2	CO1	PLO1 PLO5
CLO2	Nhận biết sự khác nhau giữa dữ liệu truyền thống và dữ liệu hiện đại. Liệt kê các phương pháp cơ bản trong khai phá dữ liệu Trình bày ứng dụng của khai phá dữ liệu.	2	CO2	PLO1
CLO3	Hiểu và mô tả được quá trình khai phá dữ liệu Mô tả một số thuật phổ biến trong khai phá dữ liệu. Sử dụng được các công cụ phổ biến trong khai phá dữ liệu. Trình bày được cách triển khai các ứng dụng khai	2	CO3	PLO1 PLO2

	phá dữ liệu.			
--	--------------	--	--	--

11.3. Ma trận đóng góp của môn học cho PLO

Mã CDR CTĐT	PLO1	PLO2	PLO5
Mã CDR MH			
CLO1	2		2
CLO2	2		
CLO3	2	2	

12. Phương pháp dạy và học

Phương pháp “Người học là trung tâm” sẽ được sử dụng trong môn học để giúp sinh viên tham gia tích cực. Kết quả học tập dự kiến sẽ đạt được thông qua một loạt các hoạt động học tập ở trường và ở nhà.

- 50% giảng dạy, 50% hướng dẫn thực hành và hỗ trợ sinh viên khi thực hành, thảo luận, đặt câu hỏi và trả lời câu hỏi.

- Tại lớp, giảng viên giải thích các định nghĩa, nền tảng lý thuyết, cách sử dụng các ứng dụng; đặt ra các vấn đề, hướng dẫn và khuyến khích sinh viên giải quyết; sau đó tóm tắt nội dung của bài học. Giảng viên cũng trình bày và thực hành làm mẫu cho sinh viên.

- Sinh viên cần lắng nghe, ghi chép và được khuyến khích nêu lên các câu hỏi, giải quyết các vấn đề và thực hành các bài tập dưới sự hướng dẫn của giảng viên.

Quy định về hình thức giảng dạy: Giảng viên có thể chủ động lựa chọn hình thức giảng dạy là trực tuyến (online) kết hợp trực tiếp (offline), đảm bảo tổng thời gian giảng dạy trực tuyến không vượt quá 30% thời gian giảng dạy của cả môn học.

13. Yêu cầu môn học

- Hoàn thành tất cả bài tập cá nhân và bài tập nhóm.
- Đọc tài liệu môn học theo yêu cầu của giảng viên.
- Thực hiện các hoạt động khác theo yêu cầu của giảng viên.

14. Học liệu của môn học

14.1. Tài liệu chính

[1] Lê Văn Quốc Anh (2018), *Giáo trình Khai phá Dữ liệu*, NXB Giao thông Vận tải.

14.2. Tài liệu tham khảo

[2] Ian H. Witten & Eibe Frank (2005), *Data mining*, Second Edition.

[3] Jiawei Han and Micheline Kamber (2012), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, Second Edition.

B. PHƯƠNG THỨC ĐÁNH GIÁ MÔN HỌC

1. Các thành phần đánh giá môn học

Thành phần đánh giá	Phương thức đánh giá	Các CĐR MH	Trọng số
A.1. Đánh giá quá trình	A.1.1. Chuyên cần, ý thức, thái độ học tập	CLO1, CLO2, CLO3	10%
	A.1.2. Kiểm tra: Bài tập cá nhân (tại lớp); Báo cáo về một thuật toán, chạy chương trình của học máy cho kế toán.	CLO1, CLO2, CLO3	20%
	A.1.3. Tiểu luận nhóm	CLO1, CLO2, CLO3	20%
A.2. Đánh giá cuối kỳ	A.2.1. Thi cuối kỳ: Kiến thức tổng quan, thuật toán và chương trình về học máy xử lý dữ liệu cho kế toán.	CLO1, CLO2, CLO3	50%

2. Nội dung và phương pháp đánh giá

2.1. Đánh giá quá trình

2.1.1. Chuyên cần

- Điểm danh ngẫu nhiên 4 buổi trong suốt khóa học.
- Đánh giá thái độ sinh viên tham gia xây dựng bài học.

2.1.2. Tiểu luận nhóm

- Làm việc trong 1 nhóm gồm từ 5 đến 7 sinh viên; nhóm sẽ chọn tùy ý một thuật toán học máy cho kế toán và được giảng viên chấp thuận.

- Nhóm sẽ trình bày chi tiết dưới dạng báo cáo tiểu luận (theo định dạng do giảng viên đưa ra) về thuật toán, chương trình học máy cho kế toán.

- Phương thức đánh giá:

- Phát biểu trình bày báo cáo (2 điểm)
- Hình thức trình bày (2 điểm)
- Thuật toán (2 điểm)
- Chạy chương trình (4 điểm)

2.1.3. Bài kiểm tra cá nhân

- Kiểm tra thực hành cá nhân là để đánh giá kiến thức và kỹ năng thực hành của mỗi sinh viên sau khi hoàn thành chương 3, 4, 5.

- Bài kiểm tra gồm 5 câu hỏi (mỗi câu 2 điểm) liên quan về các thuật toán trong học máy cho kế toán.

- Thời gian làm bài 60 phút, thực hiện sau khi kết thúc chương 5.

2.2. Thi cuối kỳ

- Đề thi được chọn ngẫu nhiên từ ngân hàng đề thi, mỗi ca thi 02 đề, mỗi đề 5 câu tự luận, thời gian làm bài 60 phút.

- Sinh viên được sử dụng tài liệu giấy, không được sử dụng Internet, không sử dụng điện thoại di động.

- Phương thức đánh giá: Được chấm 2 lượt độc lập bởi 2 giảng viên có tham gia giảng dạy môn học máy cho kế toán. Điểm bài thi được chấm theo đáp án Ngân hàng đề thi môn học máy cho kế toán, mỗi câu đúng được 2 điểm. Tổng cộng 10 điểm.

2.3. Một số rubrics đánh giá

2.3.1.1. Chuyên cần

Tiêu chí đánh giá	Trọng số	Thang điểm			
		Dưới 5	5 – dưới 7	7 – dưới 9	9 - 10
Điểm danh mỗi buổi học	75%	Số buổi vắng 4 buổi	Vắng 3 buổi	Vắng 1, 2 buổi	Không vắng buổi học nào
Phát biểu, tham gia xây dựng bài	25%	Không bao giờ phát biểu trong giờ học	Có 1 lần phát biểu đúng hoặc giải bài tập tại lớp	Có 2, 3 lần phát biểu đúng hoặc giải bài tập đúng tại lớp	Có từ 4 lần trở lên phát biểu đúng hoặc giải bài tập đúng tại lớp

2.3.1.2. Tiểu luận nhóm

Tiêu chí đánh giá	Trọng số	Thang điểm			
		Dưới 5	5 – dưới 7	7 – dưới 9	9 - 10
Hình thức và bố cục	40%	Bố cục không hợp lý. Hình thức trình bày xấu. Văn phong không mạch lạc	Bố cục cơ bản hợp lý. Hình thức trình bày chấp nhận được. Văn phong đôi chỗ chưa mạch lạc	Bố cục khá hợp lý. Hình thức trình bày khá. Văn phong mạch lạc.	Bố cục logic, khoa học. Hình thức trình bày tốt. Văn phong chặt chẽ, mạch lạc
Nội dung	40%	Thông tin và số liệu không	Thông tin và số liệu tương	Thông tin và số liệu tương	Thông tin và số liệu

		đầy đủ, không phù hợp. Đánh giá thông tin chưa chính xác. Phân tích thông tin chưa chặt chẽ	đôi đầy đủ, đôi chỗ chưa phù hợp Đánh giá thông tin chưa hoàn toàn chính xác Phân tích thông tin chưa thực sự chặt chẽ	đôi đầy đủ, phù hợp Đánh giá thông tin tương đối chính xác Phân tích thông tin tương đối chặt chẽ	đầy đủ, phù hợp Đánh giá thông tin chính xác Phân tích thông tin chặt chẽ
Phối hợp giữa các thành viên	20%	Đóng góp ý kiến: chỉ dựa vào ý kiến 1 người. Đóng góp thu thập tài liệu: chỉ 1 người cung cấp. Đóng góp viết báo cáo: chỉ 1 người viết báo cáo.	Đóng góp ý kiến: chỉ dựa vào ý kiến 2 người. Đóng góp thu thập tài liệu: 2 người cung cấp. Đóng góp viết báo cáo: 2 người viết báo cáo.	Đóng góp ý kiến: nhiều người nhưng không phải tất cả. Đóng góp thu thập tài liệu: nhiều người nhưng không phải tất cả. Đóng góp viết báo cáo: nhiều người nhưng không phải tất cả.	Đóng góp ý kiến: toàn bộ mọi người. Đóng góp thu thập tài liệu: tất cả thành viên. Đóng góp viết báo cáo: toàn bộ.

(Chú thích: Bảng này dùng để đánh giá chung cho cả nhóm)

C. NỘI DUNG CHI TIẾT GIẢNG DẠY

Thời lượng (tiết)	Nội dung giảng dạy chi tiết	CĐR của môn học	Hoạt động dạy và học	Minh chứng đánh giá	Mục tiêu chương
5 LT	CHƯƠNG 1: TỔNG QUAN VỀ KHAI PHÁ DỮ LIỆU 1.1 Khái niệm 1.2 Các bước của quá trình khai phá dữ liệu 1.3 Các loại dữ liệu cho khai phá 1.4 Độ đo trong khai phá dữ liệu 1.5 Vấn đề thu thập và chuẩn bị dữ liệu 2.2 Vấn đề chất lượng dữ liệu 2.3 Trực quan hóa dữ liệu	CLO1	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên trình bày một số khái niệm, lịch sử, chức năng và cấu trúc tổng thể; sau đó hướng dẫn sinh viên tìm kiếm thông tin và thảo luận tại chỗ về một số nội dung liên quan. SINH VIÊN: Sử dụng Internet tra cứu thông tin và thảo luận tại chỗ về kiến thức.	Kiểm tra quá trình	Nắm vững các Định nghĩa thuật ngữ liên quan đến khai phá dữ liệu. Phát biểu được các đặc trưng của khai phá dữ liệu. Trình bày được quá trình phát triển của khai phá dữ liệu, thuật toán của Python.

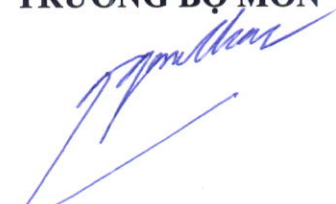
5 LT	CHƯƠNG 2:	CLO2	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên trình bày một số khái niệm, lịch sử, chức năng và cấu trúc tổng thể; sau đó hướng dẫn sinh viên tìm kiếm thông tin và thảo luận tại chỗ về một số nội dung liên quan đến dữ liệu và lịch sử phát triển của học máy. SINH VIÊN: Sử dụng Internet tra cứu thông tin và thảo luận tại chỗ về kiến thức học máy cho kế toán.	Kiểm tra quá trình, thi cuối kỳ	Nhận biết sự khác nhau giữa dữ liệu truyền thống và dữ liệu hiện đại. Liệt kê các phương pháp cơ bản trong khai phá dữ liệu
5 TH	CÁC VẤN ĐỀ TIỀN XỬ LÝ DỮ LIỆU 2.1 Tổng quan về giai đoạn tiền xử lý dữ liệu 2.2 Tóm tắt dữ liệu 2.3 Làm sạch dữ liệu 2.4 Tích hợp dữ liệu 2.5 Biến đổi dữ liệu 2.6 Thu giảm dữ liệu 2.7 Rời rạc hóa dữ liệu 2.8 Tạo cây phân cấp ý niệm	CLO2, CLO3	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên hướng dẫn nội dung	Thi cuối kỳ	Hiểu và mô tả được quá trình khai phá dữ liệu Mô tả một số
5 LT	CHƯƠNG 3: PHÂN LOẠI DỮ LIỆU				
5 TH	3.1 Tổng quan về phân loại dữ liệu 3.2 Phân loại dữ liệu với cây quyết định 3.3 Phân loại dữ liệu với mạng Bayes				

	3.4 Phân loại dữ liệu với mạng Neural 3.5 Các phương pháp phân loại dữ liệu khác		các thuật toán; sau đó, giảng viên hướng dẫn sinh viên thực hành thông qua bài tập tình huống SINH VIÊN: + Tại nhà: Ôn lại kiến thức tổng quan về học máy cho kế toán. + Tại lớp: Thực hành các nội dung về cài đặt, thuật toán của học máy cho kế toán.		thuat phổ biến trong khai phá dữ liệu. Sử dụng được các công cụ phổ biến trong khai phá dữ liệu. Trình bày được cách triển khai các ứng dụng khai phá dữ liệu.
5 LT + 5 TH	5 LT CHƯƠNG 4: GOM CỤM DỮ LIỆU 4.1 Tổng quan về gom cụm dữ liệu 4.2 Gom cụm dữ liệu bằng phân hoạch 4.3 Gom cụm dữ liệu kiểu phân cấp 4.4 Gom cụm dữ liệu dựa trên mật độ 4.5 Các phương pháp gom cụm dữ liệu khác	CLO3,	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên hướng dẫn các nội dung về thuật toán học máy cho kế toán; sau đó, giảng viên hướng dẫn sinh viên thực hành thông qua bài tập tình huống.	Kiểm tra quá trình và thi cuối kỳ	Hiểu và mô tả được quá trình khai phá dữ liệu Mô tả một số thuật phổ biến trong khai phá dữ liệu. Sử dụng được các

			SINH VIÊN: + Tại nhà: Ôn lại quy trình thiết lập thuật toán của học máy cho kế toán. + Tại lớp: Thực hành các nội dung về thuật toán học máy cho kế toán.		công cụ phổ biến trong khai phá dữ liệu. Trình bày được cách triển khai các ứng dụng khai phá dữ liệu.
5LT + 10TH	CHƯƠNG 5: KHÁM PHÁ MẪU VÀ TÌM LUẬT KẾT HỢP 5.1 Tổng quan về mẫu và khám phá mẫu 5.2 Khai thác tập phổ biến và luật kết hợp 5.3 Giải thuật Apriori 5.4 Giải thuật FP-Growth 5.5 Khám phá luật kết hợp dựa trên ràng buộc 5.6 Phân tích tương quan	CLO3	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên hướng dẫn các nội dung thuật toán học máy cho kế toán; sau đó, giảng viên hướng dẫn sinh viên thực hành thông qua bài tập tình huống. SINH VIÊN: + Tại nhà: Ôn lại và luyện tập thuật toán học máy cho kế toán. + Tại lớp: Thực hành các nội	Kiểm tra quá trình và thi cuối kỳ	Sử dụng được các công cụ phổ biến trong khai phá dữ liệu. Trình bày được cách triển khai các ứng dụng khai phá dữ liệu.

			dung về thuật toán học máy cho kế toán.		
5 LT + 5 TH	CHƯƠNG 6: KHAI THÁC DỮ LIỆU VỚI DỮ LIỆU CÓ CẤU TRÚC PHỨC TẠP 6.1 Khai thác dòng dữ liệu 6.2 Khai thác dữ liệu Time-Series 6.3 Khai thác dữ liệu dạng đồ thị và dữ liệu mạng xã hội 6.4 Khai thác dữ liệu văn bản và dữ liệu đa phương tiện	CLO1, CLO2	Trực tiếp (offline) hay trực tuyến (online) GIẢNG VIÊN: Giảng viên hướng dẫn các nội dung về thuật toán học máy cho kế toán; sau đó, giảng viên hướng dẫn sinh viên thực hành thông qua bài tập tình huống. SINH VIÊN: + Tại nhà: Ôn lại và luyện tập thuật toán học máy cho kế toán. + Tại lớp: Thực hành các nội dung về thuật toán học máy cho kế toán.	Thi cuối kỳ	Sử dụng được các công cụ phổ biến trong khai phá dữ liệu. Trình bày được cách triển khai các ứng dụng khai phá dữ liệu.
Môn học được giảng dạy trực tuyến tối đa không quá 30% tổng thời lượng chương trình.					

TRƯỞNG BỘ MÔN


PGS.TS. Nguyễn Minh Hải

NGƯỜI BIÊN SOẠN


TS. Nguyễn Minh Tùng

KT. HIỆU TRƯỞNG

PHÓ HIỆU TRƯỞNG



PGS.,TS. Nguyễn Đức Trung