TRƯỜNG ĐẠI HỌC NGÂN HÀNG
THÀNH PHỐ HỒ CHÍ MINH
BỘ MÔN TOÁN KINH TẾ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

TP. Hồ Chí Minh, ngày 28 tháng 2 năm 2024

ĐỀ CƯƠNG MÔN HỌC
KHOA HỌC DỮ LIỆU CHO KINH DOANH NÂNG CAO

A. THÔNG TIN CHUNG VỀ MÔN HỌC

1. **Tên môn học (tiếng Việt) : KHOA HỌC DỮ LIỆU CHO KINH DOANH NÂNG CAO**
2. **Tên môn học (tiếng Anh) : ADVANCED DATA SCIENCE FOR BUSINESS**
3. **Mã số môn học : DAT728**
4. **Trình độ đào tạo : Đại học**
5. **Ngành đào tạo áp dụng : Các ngành đào tạo bậc đại học**
6. **Số tín chỉ : 02**
7. **Phân bổ thời gian : 100 giờ**
   - Trực tiếp : 30 giờ
     - *Lý thuyết : 30 giờ*
     - *Thực hành : 0 giờ*
   - Trực tuyến (nếu có) : không quá 30% tổng số giờ giảng.
   - Tự học, tự nghiên cứu : 70 giờ
   - Khác : 0 giờ
8. **Khoa quản lý môn học : Bộ môn Toán Kinh tế**
9. **Môn học trước : Khoa học dữ liệu cho kinh doanh**
10. **Mô tả môn học**

Môn học trang bị trang bị cho sinh viên các kiến thức về khoa học dữ liệu ứng dụng trong kinh doanh sử dụng ngôn ngữ lập trình Python và các công cụ trong hệ sinh thái ứng dụng khoa học dữ liệu vào lĩnh vực kinh doanh. Cụ thể, các sinh viên sẽ được học lý thuyết và thực hành các cộng cụ về xử lý, thống kê và trực quan hoá dữ liệu như Matplotlib, Numpy, Numba, Pandas, SciPy, Scikit-Learn, StatsModels, và nhiều công cụ hiện đại khác được cập nhật liên tục theo sự phát triển công nghệ trong lĩnh vực

khoa học dữ liệu. Sinh viên cũng được học và thực hành với các giải thuật học máy cơ bản, bao gồm các kiến thức và giải thuật học có giám sát, học không có giám sát ứng dụng vào phân loại, thu giảm chiều và gom cụm dữ liệu. Hơn nữa, các kỹ thuật tinh chỉnh tham số và các độ đo hiệu quả của các giải thuật học cũng được giới thiệu cho sinh viên.

Học xong môn học này, sinh viên có thể vận dụng các kiến thức khoa học dữ liệu và học máy cơ bản vào phân tích dữ liệu kinh doanh thuộc nhiều mảng khác nhau, ví dụ như FinTech, quản lý tài sản, tài chính doanh nghiệp, . . và sử dụng kiến thức đã học cho các môn học tiếp theo trong chương trình đào tạo.

11. Mục tiêu và chuẩn đầu ra của môn học

11.1. Mục tiêu của môn học

Ký hiệu mục tiêu	Mô tả mục tiêu	Nội dung CĐR CTĐT phân bổ cho môn học	Ký hiệu CĐR CTĐT (PLOn)
(a)	(b)	(d)	(e)
CO1	Môn học trang bị cho sinh viên kỹ năng làm việc độc lập và làm việc theo nhóm để giải quyết các bài toán liên quan đến khoa học dữ liệu trong kinh doanh.	Có năng lực làm việc theo nhóm, truyền đạt hiệu quả trong điều kiện làm việc thay đổi và môi trường hội nhập quốc tế.	PLO3
CO2	Người học hiểu, đánh giá được các công cụ dành cho phân tích và trực quan hoá dữ liệu như Matplotlib, Numpy, Numba, Pandas, SciPy, Scikit-Learn, StatsModels, . . . để từ đó xây dựng mô hình phù hợp cho các bài toán về khoa học dư liệu trong kinh doanh.	Có khả năng vận dụng thành thạo kiến thức nền tảng và chuyên sâu để giải quyết các vấn đề chuyên môn	PLO6

Ký hiệu mục tiêu	Mô tả mục tiêu	Nội dung CĐR CTĐT phân bổ cho môn học	Ký hiệu CĐR CTĐT (PLOn)
(a)	(b)	(d)	(e)
CO3	Người học có khả năng vận dụng cao các kiến thức về khoa học dữ liệu và học máy đã học một cách hệ thống để giải quyết các vấn đề chuyên môn trong lĩnh vực kinh doanh, tài chính - ngân hàng	Có kiến thức và kỹ năng về khoa học dữ liệu ứng dụng trong kinh doanh	PLO11

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)

Ký hiệu CĐR MH (CLOi)	Nội dung CĐR MH	Mức độ theo thang đo	Đáp ứng mục tiêu môn học (COx)	Ký hiệu CĐR CTĐT (PLOn)
(a)	(b)	(c)	(d)	(e)
CLO1	Ứng dụng Pandas vào xử lý dữ liệu, biến đổi dữ liệu thô thành dạng có thể sử dụng thống kê, dự đoán (dự báo), phân loại, thu giảm chiều hoặc gom cụm.	3	CO1, CO2	PLO3
CLO2	Ứng dụng được các công cụ như Matlibplot Seaborn, . . . vào trực quan hoá dữ liệu bằng các biểu đồ khác nhau.	3	CO2	PLO3
CLO3	Ứng dụng được các mô hình thống kê và học máy sử dụng Statsmodels và sciKit-learn vào phân tích dữ liệu; PCA, t-SNE, UMAP, k-means vào	3	CO1, CO3	PLO6

Ký hiệu CĐR MH (CLOi)	Nội dung CĐR MH	Mức độ theo thang đo	Đáp ứng mục tiêu môn học (COx)	Ký hiệu CĐR CTĐT (PLOn)
(a)	(b)	(c)	(d)	(e)
	thu giảm chiều dữ liệu và gom cụm dữ liệu.			
CLO4	Ứng dụng các kiến thức về khoa học dữ liệu và học máy vào mô hình hoá các bài toán thực tế.	4	CO2, CO3	PLO11
CLO5	Xây dựng các giải pháp trực quan hoá, phân loại, dự báo, gom cụm dữ liệu ứng dụng trong lĩnh vực kinh doanh, tài chính ngân hàng.	3	CO2, CO3	PLO6

11.3. Ma trận tích hợp giữa CĐR MH (CLOi), CĐR CTĐT (PLOn) và Chỉ số đánh giá kết quả thực hiện (PIn.k):

	PLO3		PLO6		PLO11	
	PI 3.1	**PI 3.2**	**PI 6.1**	**PI 6.2**	**PI 11.1**	**PI 11.2**
CLO1	3	3				
CLO2	3	3				
CLO3			3	3		
CLO4					4	4
CLO5			3	3		

12. Phương pháp và hình thức dạy và học

12.1. Phương pháp dạy và học:

Nhằm giúp sinh viên đạt được các mục tiêu và chuẩn đầu ra của môn học, các chiến lược và phương pháp giảng dạy được áp dụng cụ thể như sau:

- Phương pháp Thuyết giảng (Lecture): Giảng viên trình bày tài liệu và giải đáp các thắc mắc của học viên.
- Phương pháp Bài giảng tương tác (Interactive lecture): Thuyết giảng trong đó có quãng thời gian nghỉ khoảng từ hai đến mười lăm phút để cho sinh viên thực hiện các hoạt động học tập (chẳng hạn như: trả lời các câu hỏi trắc nghiệm nhiều lựa

chọn, giải quyết vấn đề, bài tập so khớp và điền vào chỗ trống trong tập ghi chú bài giảng, vấn đáp về một trường hợp điển hình nhỏ, bài tập chia sẻ suy nghĩ theo cặp hoặc thảo luận nhóm nhỏ) cứ sau 12 đến 20 phút.

- Phương pháp Ôn luyện (Recitation): Học sinh trả lời các câu hỏi về kiến thức và đọc hiểu.
- Phương pháp Thảo luận có định hướng (Directed discussion): Thảo luận trong lớp theo một bộ câu hỏi ít nhiều có tính trật tự mà giảng viên đặt ra để dẫn dắt học sinh đến những nhận thức hoặc kết luận nhất định hoặc để giúp họ đạt được một kết quả học tập cụ thể.
- Phương pháp giảng dạy thông qua Làm việc/học tập theo nhóm (Group work/learning): người học thực hiện hoạt động học tập hoặc tạo ra một sản phẩm theo nhóm nhỏ từ hai đến sáu người trong hoặc ngoài lớp học, dưới sự hướng dẫn cẩn thận của giảng viên.

12.2. Hình thức dạy và học:

Giảng viên có thể chủ động lựa chọn hình thức giảng dạy là giảng dạy trực tiếp cho toàn bộ thời gian của môn học hoặc lựa chọn hình thức giảng dạy trực tuyến kết hợp trực tiếp, nhưng phải đảm bảo tổng thời gian giảng dạy trực tuyến không vượt quá 30% thời gian giảng dạy của cả môn học.

13. Quy định của môn học

- Tùy số sinh viên mà giảng viên quyết định số lượng thành viên nhóm học tập.
- Sinh viên tham dự lớp học phần phải tuân thủ quy tắc ứng xử của Nhà Trường; sinh viên phải đến lớp đúng giờ, đảm bảo thời gian học trên lớp, có thái độ nghiêm túc và chủ động, tích cực trong học tập, nghiên cứu.
- Sinh viên phải có ý thức chuẩn bị giáo trình, học liệu, máy tính cá nhân (khi cần thiết) để phục vụ quá trình học tập.

14. Tài liệu học tập

14.1. Tài liệu bắt buộc

[1] Rogel-Salazar, Jesús, (2020). Advanced data science and analytics with Python. Boca Raton: CRC Press, Taylor & Francis Group.

14.2. Tài liệu tham khảo

[2] Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel, (2020). Data mining for business analytics: concepts, techniques and applications in Python Wiley & Sons, Inc.

[3] Nguyễn, Ngọc Thạch, (2021). Data science for financial econometrics. Springer.

B. PHƯƠNG THỨC ĐÁNH GIÁ MÔN HỌC

1. Cơ cấu phương thức đánh giá môn học

Loại hình đánh giá	Phương pháp đánh giá	Số lượng bài đánh giá	CĐR MH được đánh giá	Trọng số
1. Đánh giá quá trình	1.1. Đánh giá tính chuyên cần			10%
	1.2. Đánh giá quá trình 1: Kiểm tra trên phòng máy		CLO1, CLO2, CLO3	20%
	1.3. Đánh giá quá trình 2: Bài tập lớn và bài tập về nhà		CLO1, CLO2, CLO3, CLO4, CLO5	20%
2. Đánh giá cuối kỳ	2.1. Đánh giá cuối kỳ		CLO1, CLO2, CLO3, CLO4	50%

2. Phương pháp và công cụ đánh giá

2.1. Đánh giá quá trình

2.1.1. Đánh giá tính chuyên cần

a. Hình thức đánh giá: sinh viên phát biểu, tự luận hoặc trắc nghiệm, trên lớp học hoặc trên các nền tảng học tập trực tuyến.

b. Thời điểm, cách thức thực hiện bài đánh giá:

Thời điểm: diễn ra trong suốt các buổi học của học phần.

Việc đánh giá kết quả học tập được thực hiện bằng đánh giá chuyên cần được thực hiện bằng phương thức điểm danh và ghi nhận quá trình tham gia học tập các nội dung của học phần. Điểm danh thực hiện trên danh sách lớp học phần chính thức do Trường cung cấp. Việc ghi nhận quá trình tham gia học tập các nội dung của học phần được thực hiện khi: (1) giảng viên mời đích danh sinh viên trả lời các câu hỏi hay phân công giải quyết các bài tập/chủ đề thảo luận (tham gia thụ động), (2) sinh viên tự giác (xung phong) trả lời các câu hỏi hay tham gia giải quyết các bài tập/chủ đề thảo luận (tham gia chủ động); sinh viên có tần suất tham gia trên 50% số buổi học với đa số lời đáp sát đáp án của vấn đề thì được xác định là chủ động tham gia rất tích cực vào quá trình học tập trên giảng đường.

c. Nội dung, kết cấu bài đánh giá

Nội dung của đánh giá chuyên cần gồm:

- tần suất hiện diện của sinh viên

- sự tham gia vào quá trình học tập tại giảng đường.

Kết cấu bài đánh giá: điểm danh, ghi nhận sự tham gia vào quá trình học.

d. Rubric đánh giá

Tiêu chí đánh giá	Trọng số	Thang điểm				
		Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Tần suất hiện diện của sinh viên	40%	Hiện diện giảng đường dưới 40% số buổi học	Hiện diện giảng đường trên 50% số buổi học	Hiện diện giảng đường trên 60% số buổi học	Hiện diện giảng đường trên 70% số buổi học	Hiện diện giảng đường trên 80% số buổi học
Sự tham gia vào quá trình học tập tại giảng đường	60%	không tham gia vào quá trình thảo luận về nội dung bài học trong suốt học phần	tham gia thụ động vào quá trình thảo luận về nội dung bài học trong suốt học phần	tham gia chủ động vào quá trình thảo luận về nội dung bài học trong suốt học phần	chủ động tham gia tích cực vào quá trình thảo luận về nội dung bài học trong suốt học phần	chủ động tham gia rất tích cực vào quá trình thảo luận về nội dung bài học trong suốt thời học phần

2.1.2. Đánh giá quá trình 1: Bài kiểm tra trên phòng máy

a. Hình thức đánh giá: sinh viên làm bài tập cá nhân lập trình trên phòng thực hành hoặc trên các nền tảng học tập trực tuyến.

b. Thời điểm, cách thức thực hiện bài đánh giá:

Thời điểm: một hoặc một phần buổi học được giảng viên và sinh viên thống nhất.

Đánh giá kết quả học tập bằng Bài kiểm tra cá nhân được thực hiện bằng cách tổ chức kiểm tra tập trung tại phòng thực hành có máy tính hỗ trợ làm bài: lập

trình trên máy và được sử dụng tài liệu (trong trường hợp cần thiết, có thể thay thế bằng hình thức kiểm tra online; trong trường hợp này, giảng viên sẽ thông báo chi tiết đến sinh viên ít nhất 1 tuần trước ngày kiểm tra về thời gian giao đề, thời gian thực hiện bài kiểm tra, phương thức làm bài và nộp bài qua mạng internet).

c. Nội dung, kết cấu bài đánh giá

Nội dung đánh giá của Bài kiểm tra cá nhân là khối lượng kiến thức kiểm tra tương ứng với khối lượng kiến thức của tiến độ dạy học đã được quy định. Đề kiểm tra do giảng viên soạn, chịu trách nhiệm về chuyên môn;

Kết cấu bài đánh giá: tối thiểu có 02 câu hỏi; thời gian kiểm tra tối đa bằng thời gian thi hết học phần.

d. Rubric đánh giá

Tiêu chí đánh giá	CĐR	Trọng số	Thang điểm				
			Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Gọi đúng thư viện để thao tác trên dữ liệu dạng bảng	CLO1	20%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Sử dụng đúng các thư viện hỗ trợ trực quan hoá dữ liệu	CLO2	40%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Sử dụng đúng các thư viện hỗ trợ các mô hình học máy cơ bản	CLO3	40%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case

2.1.3. Đánh giá quá trình 2: Bài tập lớn và bài tập về nhà

xét việc tăng thành viên của nhóm). Quy cách của báo cáo được giảng viên quy định trong 03 buổi học đầu tiên.

b. Thời điểm, cách thức thực hiện bài đánh giá:

Thời điểm: giảng viên và sinh viên thống nhất chọn thời điểm làm bài và nộp bài. Giảng viên tổ chức cho sinh viên hình thành các nhóm bài tập lớn. Đề tài của bài tập lớn, thời gian và phương thức nộp được giảng viên thông báo nhóm sinh viên trực tiếp tại lớp hoặc qua email trong tuần học đầu tiên của học phần. Nhóm sinh viên tổ chức thực hiện bài tập trong thời gian tự học tại nhà; sau đó nộp bài tập đến giảng viên theo thời gian quy định. Giảng viên chấm, trả điểm và cung cấp nhận xét cho các nhóm rút kinh nghiệm, đúc kết tri thức cần lĩnh hội.

c. Nội dung, kết cấu bài đánh giá

Nội dung đánh giá của hình thức đánh giá kết quả học tập bằng bài tập lớn theo nhóm và bài tập về nhà thực hiện bởi từng cá nhân, là khối lượng kiến thức của học phần đã quy định trong Đề cương này, thể hiện cụ thể qua các chủ đề của các bài tập

Kết cấu bài đánh giá: tối thiểu có 03 yêu cầu đối với một đề tài bài tập nhóm, tối thiểu 02 câu hỏi đối với bài tập về nhà.

d. Rubric đánh giá

Bài tập về nhà:

Tiêu chí đánh giá	CĐR	Trọng số	Thang điểm				
			Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Gọi đúng thư viện để thao tác trên dữ liệu dạng bảng	CLO1	30%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Sử dụng đúng các thư viện hỗ trợ trực quan hoá dữ liệu	CLO2	20%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case

Sử dụng đúng các thư viện hỗ trợ các mô hình học máy cơ bản	CLO3	20%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Ứng dụng các kiến thức về khoa học dữ liệu và học máy vào mô hình hoá các bài toán thực tế.	CLO4	20%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Xây dựng các giải pháp trực quan hoá và gom cụm dữ liệu trong kinh doanh, tài chính ngân hàng.	CLO5	10%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case

Bài tập lớn:

Tiêu chí đánh giá	CĐR	Trọng số	Thang điểm				
			Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Phần lập trình	CLO2	50%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Cơ sở lý luận của báo cáo	CLO3	30%	Không sử dụng lý luận khoa học liên quan	Có sử dụng nhưng chưa đúng lý luận	Trình bày lý luận khoa học liên quan,	Sử dụng đúng lý luận khoa học liên quan, có sức	Sử dụng chính xác, rất thuyết phục về

				khoa học liên quan	nhưng chưa đủ sức thuyết phục	thuyết phục	lý luận khoa học liên quan
Văn phong khoa học	CLO4	10%	Hành văn tối nghĩa ở các phần nội dung chính	Hành văn lủng củng khiến cho người đọc khó có thể hiểu nội dung	Hành văn lủng củng nhưng người đọc vẫn có thể hiểu nội dung	Hành văn tốt, đôi khi có lỗi diễn đạt	Hành văn rõ ràng, mạch lạc
Lỗi đạo văn		Trừ 100% điểm đạt được					
Hình thức của báo cáo	CLO5	10%	Không định dạng theo bất cứ tiêu chí nào	Có lỗi: không căn lề, không thống nhất định dạng đoạn văn, sai khổ giấy	Có lỗi: Không định dạng toàn văn bản, không thống nhất font chữ	Có lỗi: thiếu đánh số trang; thiếu bìa hoặc trình bày bìa sai quy định	Định dạng đúng tất cả các tiêu chí yêu cầu

2.2. Đánh giá cuối kỳ

- Đề thi do Khoa/Bộ môn quản lý học phần ra đề hoặc trích xuất từ Ngân hàng câu hỏi thi/Ngân hàng đề thi của trường (nếu có)
- Hình thức đánh giá:

Hình thức đánh giá	Lựa chọn
Vấn đáp	
Tự luận	
Trắc nghiệm khách quan	
Trắc nghiệm khách quan kết hợp tự luận	

Hình thức đánh giá		Lựa chọn
Thực hành trên máy tính		
Tiểu luận/đồ án/bài tập lớn (cá nhân)	Nộp bài không thuyết trình	
	Nộp bài có thuyết trình	
Tiểu luận/đồ án/bài tập lớn (nhóm)	Nộp bài không thuyết trình	X
	Nộp bài có thuyết trình	

- Thời gian làm bài (tối thiểu – tối đa): từ 1 đến 7 ngày làm việc.
- Mô tả về kết cấu đề thi: đề tài dự án khoa học dữ liệu trong kinh doanh: Đề tài có ít nhất 03 yêu về việc vận dụng các kiến thức đã học để giải quyết một bài toán thực tế. Sinh viên làm dự án theo nhóm. Nội dung yêu cầu của dự án phải phản ánh khía cạnh thức tế của việc ứng dụng khoa học dữ liệu trong kinh doanh. Tuỳ sỉ số lớp mà giảng viên quyết định mỗi nhóm bao nhiêu sinh viên.
- Mô tả về phạm vi nội dung của đề thi: chương 2 đến chương 6.
- Rubric đánh giá/hoặc Ma trận đề: Đánh giá tương tự như phần 2.1.3 (Bài tập lớn).

Tiêu chí đánh giá	CĐR	Trọng số	Thang điểm				
			Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Phần lập trình	CLO2	50%	Có nộp kết quả làm bài tập	Chương trình chạy không có lỗi	Chương trình chạy đúng 1/3 số test	Chương trình chạy đúng 2/3 số test case	Chương trình chạy đúng tất cả test case
Cơ sở lý luận của báo cáo	CLO3	30%	Không sử dụng lý luận khoa học liên quan	Có sử dụng nhưng chưa đúng lý luận khoa học liên quan	Trình bày lý luận khoa học liên quan, nhưng chưa đủ sức thuyết phục	Sử dụng đúng lý luận khoa học liên quan, có sức thuyết phục	Sử dụng chính xác, rất thuyết phục về lý luận khoa học liên quan

Tiêu chí đánh giá	CĐR	Trọng số	Thang điểm				
			Từ 0 đến 3,9	Từ 4,0 đến 5,4	Từ 5,5 đến 6,9	Từ 7,0 đến 8,4	Từ 8,5 đến 10
Văn phong khoa học	CLO4	10%	Hành văn tối nghĩa ở các phần nội dung chính	Hành văn lủng củng khiến cho người đọc khó có thể hiểu nội dung	Hành văn lủng củng nhưng người đọc vẫn có thể hiểu nội dung	Hành văn tốt, đôi khi có lỗi diễn đạt	Hành văn rõ ràng, mạch lạc
Lỗi đạo văn		Trừ 100% điểm đạt được					
Hình thức của báo cáo	CLO5	10%	Không định dạng theo bất cứ tiêu chí nào	Có lỗi: không căn lề, không thống nhất định dạng đoạn văn, sai khổ giấy	Có lỗi: Không định dạng toàn văn bản, không thống nhất font chữ	Có lỗi: thiếu đánh số trang; thiếu bìa hoặc trình bày bìa sai quy định	Định dạng đúng tất cả các tiêu chí yêu cầu

- Quy định về việc sử dụng tài liệu: sinh viên được sử dụng tài liệu nhưng cần ghi rõ nguồn trích dẫn/tham khảo.
- Các quy định khác về bài đánh giá: Không có.

C. NỘI DUNG CHI TIẾT GIẢNG DẠY

Thời lượng (giờ)	Nội dung giảng dạy	CĐR MH	Hoạt động Phương pháp dạy và học	Bài đánh giá	Học liệu
(a)	(b)	(c)	(d)	(e)	(f)
5	**CHƯƠNG 1: GIỚI THIỆU VỀ KHOA HỌC DỮ LIỆU** 1. Phân tích dự đoán 2. Khai phá dữ liệu và học máy 3. Học máy trong phân tích dữ liệu 4. Cách thức các giải thuật học máy hoạt động 5. Những vấn đề thường gặp với học máy 6. Vòng đời của một dự án phân tích dữ liệu: CRISP-DM 7. Dữ liệu lớn và khoa học dữ liệu 8. Ra quyết định dựa trên dữ liệu 9. Tổng kết	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực truyến (online) **GIẢNG VIÊN:** - Giới thiệu môn học (chú ý: giáo trình, công bố lịch kiểm tra giữa kỳ, giao đề tài bài tiểu luận nhóm). - Giảng bài, vấn đáp về các khái niệm cơ bản. - Giao sinh viên chuẩn bị bài Chương 2. - Giao bài tập về nhà **SINH VIÊN:** - Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên.	B.1.1 B.1.2 B.1.3 B.2.1	[1]
5	**CHƯƠNG 2: MÔ HÌNH HỒI QUY** 1. Hồi quy tuyến tính 2. Case study 1: xây dựng thuật toán hồi quy tuyến tính cho tập dữ liệu thực tế (ví dụ giá nhà ở thành phô Boston) 3. Hồi quy logistic 4. Case study 2: xây dựng thuật toán hồi quy logistic cho tập dữ	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực truyến (online) **GIẢNG VIÊN:** - Kiểm tra bài cũ. - Giảng bài mới. - Giao sinh viên chuẩn bị bài Chương 3. - Giải đáp các câu hỏi của sinh viên. - Hướng dẫn sinh viên tự học, tự nghiên cứu. - Giao bài tập về nhà và bài tập lớn - Hướng dẫn làm bài tập lớn **SINH VIÊN:**	B.1.1 B.1.2 B.1.3 B.2.1	[1]

Thời lượng (giờ)	Nội dung giảng dạy	CĐR MH	Hoạt động Phương pháp dạy và học	Bài đánh giá	Học liệu
	dataset) 5. Bài tập 6. Tổng kết Các nội dung sinh viên tự nghiên cứu: - Hồi quy tuyến tính đa biến - Hồi quy logistic đa biến - Ứng dụng logistic cho bài toán phân lớp đa lớp		- Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên.		
5	**CHƯƠNG 3: QUÁ KHỚP (OVERFITTING) & TƯ DUY PHÂN TÍCH DỮ LIỆU** 1. Quá khớp (overfitting) 2. Các kỹ thuật khắc phục overfitting 3. Giới thiệu về tư duy phân tích 4. Những thang đo cho các mô hình hồi quy 5. Những thang đo cho các mô hình phân loại 6. Khung giá trị kỳ vọng 7. Bài tập 8. Tổng kết Các nội dung sinh viên tự nghiên cứu: - Các kỹ thuật tiền xử lý dữ liệu: o chuẩn hoá min-max, chuẩn hoá softmax	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực truyến (online) **GIẢNG VIÊN:** - Kiểm tra bài cũ. - Giảng bài mới. - Giao sinh viên chuẩn bị bài Chương 4. - Giải đáp các câu hỏi của sinh viên. - Hướng dẫn sinh viên tự học, tự nghiên cứu. - Giao bài tập về nhà và bài tập lớn - Hướng dẫn làm bài tập lớn **SINH VIÊN:** - Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên.	B.1.1 B.1.2 B.1.3 B.2.1	[1]

Thời lượng (giờ)	Nội dung giảng dạy	CĐR MH	Hoạt động Phương pháp dạy và học	Bài đánh giá	Học liệu
	○ làm đầy dữ liệu ○ xử lý các điểm dữ liệu bất bình thường (outliers) - Các kỹ thuật tăng cường dữ liệu				
5	**CHƯƠNG 4: BÀI TOÁN PHÂN LOẠI (CLASSIFICATION)** 1. Ôn tập về xác suất 2. Thang đo đồng dạng (similarity) 3. *k*-láng giềng gần nhất 4. Thuật toán Naïve Bayes 5. Ước lượng hợp lý cực đại 6. Case study 1: Áp dụng thuật toán phân lớp trên tập dữ liệu thực tế (ví dụ Irish Flowers) 7. Học cây quyết định 8. Case study 2: Xây dựng thuật toán phân lớp cho tập dữ liệu thực tế (ví dụ Banking dataset) 9. Bài tập 10. Tổng kết Các nội dung sinh viên tự nghiên cứu: - Các thuật toán khác nếu bỏ giả thuyết "Naïve" trong tên thuật toán Naive Bayes - Các độ đo tương tự phổ biến được dùng trong học máy	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực tuyến (online) **GIẢNG VIÊN:** - Kiểm tra bài cũ. - Giảng bài mới. - Giao sinh viên chuẩn bị bài Chương 5. - Giải đáp các câu hỏi của sinh viên. - Hướng dẫn sinh viên tự học, tự nghiên cứu. - Giao bài tập về nhà và bài tập lớn - Hướng dẫn làm bài tập lớn **SINH VIÊN:** - Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên.	B.1.1 B.1.2 B.1.3 B.2.1	[1]

Thời lượng (giờ)	Nội dung giảng dạy	CĐR MH	Hoạt động Phương pháp dạy và học	Bài đánh giá	Học liệu
5	**CHƯƠNG 5: BÀI TOÁN PHÂN CỤM DỮ LIỆU (CLUSTERING)** 1. Gom cụm dữ liệu 2. Phân cụm *k*-means 3. Case study: xây dựng thuật toán gom cụm cho tập dữ liệu thực tế (ví dụ phân cụm khách hàng của một siêu thị) 4. Bài tập 5. Tổng kết Các nội dung sinh viên tự nghiên cứu: - Các kỹ thuật thu giảm số chiều dữ liệu: o LDA o PCA o t-SNE o UMAP - Ứng dụng thu giảm số chiều trong trực quan hoá dữ liệu và minh hoạ cho bài toán phân cụm.	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực tuyến (online) **GIẢNG VIÊN:** - Kiểm tra bài cũ. - Giảng bài mới. - Giao sinh viên chuẩn bị bài Chương 6. - Giải đáp các câu hỏi của sinh viên. - Hướng dẫn sinh viên tự học, tự nghiên cứu. - Giao bài tập về nhà và bài tập lớn - Hướng dẫn làm bài tập lớn **SINH VIÊN:** - Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên.	B.1.1 B.1.2 B.1.3 B.2.1	[1]
5	**CHƯƠNG 6: KHAI PHÁ DỮ LIỆU VĂN BẢN (TEXT MINING)** 1. Biểu diễn vector đặc trưng cho dữ liệu văn bản: a. TF-IDF b. Bag of Words	CLO1 CLO2 CLO3 CLO4 CLO5	Trực tiếp (offline) hay trực tuyến (online) **GIẢNG VIÊN:** - Kiểm tra bài cũ. - Giảng bài mới. - Hướng dẫn sinh viên ôn tập. - Thông báo một phần điểm quá trình	B.1.1 B.1.2 B.1.3 B.2.1	[1]

Thời lượng (giờ)	Nội dung giảng dạy	CĐR MH	Hoạt động Phương pháp dạy và học	Bài đánh giá	Học liệu
	c. Word Embedding 2. Khai phá dữ liệu văn bản 3. Bài tập 4. Tổng kết Các nội dung sinh viên tự nghiên cứu: - Mô hình ngôn ngữ và ứng dụng trong: o dịch máy o hệ thống hỏi đáp - Bài toán phân tích cảm xúc bình luận (sentiment analysis).		- Hướng dẫn làm bài tập lớn **SINH VIÊN:** - Nghe giảng, tham gia thảo luận, phát biểu ý kiến xây dựng bài. - Trả lời câu hỏi của giảng viên. - Xem điểm, đề nghị điều chỉnh sai sót (nếu có) và xác nhận điểm. - Trao đổi với giảng viên, cả lớp về các nội dung trong học phần còn cần được giải đáp.		

TRƯỞNG BỘ MÔN

Nguyễn Minh Hải

NGƯỜI BIÊN SOẠN

Võ Đức Vĩnh

TRƯỞNG KHOA/BỘ MÔN

Nguyễn Minh Hải

HIỆU TRƯỞNG

TRƯỜNG ĐẠI HỌC NGÂN HÀNG THÀNH PHỐ HỒ CHÍ MINH

Nguyễn Đức Trung