BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH
HCMUTE

INFO UTE LIBRARY

ĐỀ CƯƠNG CHI TIẾT MÔN HỌC

CÁC ỨNG DỤNG CỦA DỮ LIỆU LỚN - HỌC MÁY Ở QUY MÔ LỚN

BIG DATA APPLICATIONS - MACHINE LEARNING AT SCALE

(Ngành Kỹ thuật Dữ liệu - Chương trình đào tạo đại học 132TC)

HCMUTE.EDU.VN - THÁNG 9 NĂM 2019

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT
THÀNH PHỐ HỒ CHÍ MINH
----------***----------

INFO UTE LIBRARY

HCMUTE

ĐỀ CƯƠNG CHI TIẾT MÔN HỌC

CÁC ỨNG DỤNG CỦA DỮ LIỆU LỚN - HỌC MÁY Ở QUY MÔ LỚN

BIG DATA APPLICATIONS - MACHINE LEARNING AT SCALE

(Ngành Kỹ thuật Dữ liệu - Chương trình đào tạo đại học 132TC)

HCMUTE.EDU.VN – THÁNG 9 NĂM 2019

MỤC LỤC

LỜI NÓI ĐẦU 4
GIỚI THIỆU CÁC DỊCH VỤ UTE Library 5
GIỚI THIỆU CÁC CƠ SỞ DỮ LIỆU TRỰC TUYẾN 9
ĐỀ CƯƠNG CHI TIẾT MÔN HỌC 10
GIỚI THIỆU CÁC CHƯƠNG TRÌNH ĐÀO TẠO 19

LỜI NÓI ĐẦU

Công nghệ thông tin (IT) là một trong những động lực quan trọng bậc nhất của sự phát triển, cùng với một số ngành công nghệ cao khác đang làm biến đổi sâu sắc đời sống kinh tế, văn hóa, xã hội của thế giới hiện đại. Việc ứng dụng và phát triển CNTT ở nước ta đã góp phần giải phóng sức mạnh vật chất, trí tuệ và tinh thần, thúc đẩy công cuộc đổi mới, phát triển nhanh và hiện đại hóa các ngành kinh tế, tăng cường năng lực cạnh tranh của các doanh nghiệp, hỗ trợ có hiệu quả cho quá trình chủ động hội nhập kinh tế quốc tế, nâng cao chất lượng cuộc sống của nhân dân, đảm bảo an ninh quốc phòng góp phần thực hiện thắng lợi sự nghiệp công nghiệp hóa, hiện đại hóa đất nước.

Hiện nay, ngành CNTT là một trong những ngành học được chú trọng trong hệ thống đào tạo của các trường Đại học trong đó có trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh. Nó được xem là ngành đào tạo mũi nhọn hướng đến sự phát triển của công nghệ và khoa học kỹ thuật trong thời đại số hóa.

Chương trình đào tạo ngành CNTT trang bị cho người học những kiến thức nền tảng và chuyên sâu về lĩnh vực công nghệ thông tin để nâng cao tay nghề nhằm phát triển khả năng sửa chữa, xây dựng, cài đặt, bảo trì các phần cứng của máy tính cũng như nghiên cứu và phát triển các ứng dụng phần mềm. Ngoài ra cũng được trang bị kiến thức về an toàn và bảo mật thông tin mạng, một trong những lĩnh vực quan trọng được quan tâm hàng đầu trên thế giới hiện nay.

Để đáp ứng nhu cầu tìm kiếm và sử dụng tài liệu học tập các môn học chuyên ngành "Công nghệ thông tin" của sinh viên. Thư viện Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh đã biên soạn tài liệu thông tin về đề cương chi tiết môn học nhằm cung cấp cho sinh viên có thể nghiên cứu, tìm hiểu, học tập về các môn học chuyên ngành.

Tài liệu thông tin về đề cương chi tiết môn học ngành "Công nghệ thông tin" nhằm hướng dẫn tìm kiếm và sử dụng tài liệu học tập các môn học chuyên ngành "Công nghệ thông tin" sẽ mang đến cho sinh viên nắm được phương pháp học nhanh nhất và đạt hiệu quả.

Mặc dù chúng tôi đã cố gắng, song việc biên soạn tài liệu chắc chắn không tránh khỏi thiếu sót, rất mong nhận được những ý kiến đóng góp từ phía người sử dụng để lần biên soạn sau được hoàn chỉnh hơn.

Thư viện ĐHSPKT TP.HCM
028.389 69 920
thuvien@hcmute.edu.vn
thuvienspkt@hcmute.edu.vn
facebook.com/hcmute.lib

GIỚI THIỆU CÁC DỊCH VỤ UTE LIBRARY

Đồng hành cùng độc giả trên con đường chinh phục tri thức

Nơi cung cấp nguồn lực thông tin khoa học kỹ thuật và giáo dục phục vụ cho nhu cầu đào tạo của Nhà trường.

Cung cấp thông tin

- ✓ Nội dung phong phú
- ✓ Đa dạng loại hình
- ✓ Cập nhật thường xuyên

Hình thức phục vụ

- ✓ Đọc tại chỗ
- ✓ Mượn về nhà
- ✓ Khai thác tài nguyên số 24/24
- ✓ Các dịch vụ học tập trực tuyến

Các loại hình dịch vụ

1. Phục vụ trực tiếp tại thư viện: được hướng dẫn tận tình với hệ thống phòng đọc & Giáo trình mở, có thể tìm đọc tài liệu dạng giấy, CD-ROM, CSDL trực tuyến,…
2. Hỗ trợ công tác biên soạn xuất bản giáo trình và tài liệu học tập, kỷ yếu hội thảo (phục vụ hoạt động NCKH, chương trình đào tạo 150 tín chỉ, bổ sung hồ sơ xét các chức danh khoa học,...).
3. Thiết kế website phục vụ hoạt động học tập, nghiên cứu khoa học, hội thảo: Xây dựng template chung, Thiết lập các trang con (sub pages), Cấp quyền và chuyển giao quyền quản trị trang con.
4. Xuất bản kỷ yếu hội thảo: Tư vấn, thiết kế, dàn trang, Thiết kế các hình ảnh, nhãn hiệu liên quan đến hội nghị (logo hội nghị, banner, poster…), Đăng ký và xin giấy phép xuất bản kỷ yếu, Giám sát các tài liệu liên quan đến chương trình như thư, thông tin hội nghị, tài liệu tham khảo,…
5. Xây dựng các sản phẩm phục vụ hội thảo và các hoạt động học tập: CD-ROM chương trình và kỷ yếu, Ứng dụng tự chạy giới thiệu tài liệu, ghi đĩa CD- ROM, DVD, USB,…
6. Cung cấp thông tin theo yêu cầu (danh mục tài liệu, tài liệu chuyên ngành, kết quả nghiên cứu khoa học, tiêu chuẩn, phát minh sáng chế…).
7. Mô tả, tạo và chọn danh mục “Tài liệu tham khảo” cho luận văn, luận án, đề tài nghiên cứu khoa học theo chuẩn quốc tế.
8. Chuyển dạng tài liệu (từ tài liệu dạng giấy sang file PDF, từ file PDF sang file Word).
9. Cung cấp các thiết bị hỗ trợ cho việc học tập và nghiên cứu,...
10. Hỗ trợ, tư vấn và phối hợp với Thư viện các trường trong việc phát triển các sản phẩm đặc thù của từng đơn vị.

Với nhiều hình thức phục vụ phong phú, thuận tiện cho người học

NGUỒN TÀI NGUYÊN THÔNG TIN

- CSDL Giáo trình và Tài liệu học tập
- CSDL Sách tham khảo Việt văn
- CSDL Luận văn, Luận án
- CSDL Sách tham khảo Ngoại văn
- CSDL các bài báo khoa học (các bài báo nghiên cứu của học viên theo chương trình đào tạo sau đại học được đăng trên các tạp chí chuyên ngành)

Địa chỉ liên hệ:

Thư viện Trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh, Số 1-3 Võ Văn Ngân, Phường Linh chiểu, Quận Thủ Đức, Tp. Hồ Chí Minh.

Điện thoại: (+84 028) 3896 9920, 3721223 (nội bộ 8223)

Email: thuvienspkt@hcmute.edu.vn

http://thuvien.hcmute.edu.vn

http://thuvienso.hcmute.edu.vn

GIỚI THIỆU CÁC WEBSITE PHÁT HÀNH UTE EBOOK

"UTE EBOOK" là những tác phẩm chỉ có thể dùng các công cụ điện tử như máy vi tính, máy trợ giúp kỹ thuật số cá nhân (thiết bị đọc, máy tính bảng và điện thoại thông minh như iPhone, iPad, Samsung Galaxy, HTC Tablet,...) để xem, đọc, và truyền tải.

"UTE EBOOK" là hệ thống phân phối sách điện tử, cung cấp đến người dùng các nội dung sách,giáo trình, truyện, tạp chí, tài liệu, chuyên đề và các thể loại văn bản đọc, audio khác trên di động, thông qua các kênh website, wapsite và client giúp khách hàng có thể cập nhật thông tin kiến thức mọi lúc mọi nơi.

Stt	Tên đơn vị phát hành	Website	Truy cập nhanh kho giáo trình
1	Nhà Xuất Bản Tổng Hợp Thành Phố Hồ Chí Minh	https://sachweb.com	https://bit.ly/2LSRzXU
2	Công Ty Cổ Phần Tin Học Lạc Việt	http://sachbaovn.vn	https://bit.ly/2Zx8YZn
3	Công Ty TNHH Sách Điện Tử Trẻ (YBOOK)	http://ybook.vn/	https://bit.ly/2GHF2lQ
4	Công Ty Cổ Phần Dịch Vụ Trực Tuyến	https://read.alezaa.com	https://goo.gl/4MM7RM

	VINAPO		
5	Công Ty Cổ Phần Thương Mại Dịch Vụ Mê Kông COM	http://reader.vinabook.com	https://goo.gl/i6Qpb1
6	Thư viện Trường Đại học Sư phạm Kỹ thuật TP. HCM	http://thuvien.hcmute.edu.vn/	http://thuvien.hcmute.edu.vn/

DỊCH VỤ HỖ TRỢ XUẤT BẢN ĐIỆN TỬ:
Sách chuyên khảo, giáo trình, sách tham khảo, sách hướng dẫn, tài liệu phục vụ đào tạo, Kỷ yếu hội thảo, tập san, tạp chí

Nhằm hỗ trợ Quý Thư viện các Trường Đại học, Cao đẳng về việc xuất bản giáo trình, tài liệu học tập điện tử nội sinh,…đạt chất lượng cao phục vụ hoạt động đào tạo, giảng dạy; Thư viện Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh giới thiệu đến Quý Thư viện các Trường Đại học, Cao đẳng về "*Chương trình hỗ trợ xuất bản điện tử: Sách chuyên khảo, giáo trình, sách tham khảo, sách hướng dẫn, tài liệu phục vụ đào tạo, Kỷ yếu hội thảo, tập san, tạp chí* ".

1. Cơ sở xây dựng chương trình:

Căn cứ theo công văn Số: 4301/BGDĐT-GDTX ngày 20 tháng 9 năm 2019 của Bộ trưởng Bộ Giáo dục và Đào tạo về việc xây dựng và phát triển tài nguyên giáo dục mở tại website: https://itrithuc.vn

2. Thời gian triển khai chương trình hỗ trợ

- Giai đoạn 05 năm (Từ ngày 01/10/2019 đến hết ngày 01/10/2024)

3. Tổ chức chương trình

- Thư viện Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh.

4. Đơn vị công bố phổ biến xuất bản phẩm điện tử

- Thư viện Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh & Các đơn vị liên kết.
- website: http://thuvien.hcmute.edu.vn
- website: https://itrithuc.vn
- website các đơn vị liên kết.

5. Cấu trúc của xuất bản điện tử giáo trình, tài liệu học tập nội sinh

Xuất bản điện tử giáo trình, tài liệu học tập nội sinh có cấu trúc như sau:

- Trang bìa.
- Trang bìa phụ.
- Lời nói đầu: Trình bày đối tượng sử dụng, mục đích yêu cầu khi sử dụng, cấu trúc nội dung, điểm mới của giáo trình, hướng dẫn cách sử dụng, phân công nhiệm vụ của tác giả biên soạn.

- Mục lục.
- Danh mục các từ viết tắt (nếu có).
- Danh mục hình (nếu có)..
- Danh mục bảng biểu (nếu có)..
- Nội dung chính: Trình bày các chương, mục, tiểu mục và nội dung chi tiết của từng chương, mục, tiểu mục, nội dung thảo luận xêmina, câu hỏi ôn tập, bài tập, các nhiệm vụ tự học và tài liệu học tập từng chương.
- Phụ lục (nếu có).
- Tài liệu tham khảo.

6. Để biết thêm thông tin chi tiết Quý Thầy/Cô liên hệ

- Thư viện Trường Đại Học Sư phạm Kỹ thuật TP. Hồ Chí Minh.
- Số 1-3 Võ Văn Ngân, Phường Linh Chiểu, Quận Thủ Đức, TP. Hồ Chí Minh.
- Email: luatvt@hcmute.edu.vn; thuvienspkt@hcmute.edu.vn
- Điện thoại: 08.38969920 hoặc 0909836920; 0906836920 (ThS. Vũ Trọng Luật)

Xin trân trọng giới thiệu đến Quý Thầy/Cô!

GIỚI THIỆU CÁC CƠ SỞ DỮ LIỆU TRỰC TUYẾN TÌM KIẾM VÀ SỬ DỤNG HIỆU QUẢ

Thư viện ĐH SPKT TP. HCM
http://thuvien.hcmute.edu.vn
thuvienspkt@hcmute.edu.vn
028.38969920

Nhằm tăng cường năng lực nghiên cứu của tập thể giảng viên và sinh viên của nhà trường, từng bước nâng cao chất lượng đào tạo và nâng vị thế của trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh lên tầm cao mới, xứng đáng là một trong những trường đại học hàng đầu của khu vực, nhà trường đã tiến hành mua quyền sử dụng tài khoản truy cập các cơ sở dữ liệu trực tuyến để giảng viên và sinh viên của trường khai thác. Việc khai thác các tư liệu điện tử, các CSDL trực tuyến sẽ góp phần tạo thêm nhiều cơ hội tiếp cận với các kho tư liệu học thuật và nghiên cứu có giá trị cao và được phổ biến trên toàn thế giới.

Để triển khai sử dụng và khai thác các cơ sở dữ liệu trực tuyến đạt hiệu quả, Thư viện xin thông báo đến Quý Thầy Cô & Các Bạn Sinh viên về nội dung triển khai và sử dụng CSDL như sau:

Thông tin triển khai sử dụng:

- Truy cập Web, không giới hạn số lượng người sử dụng, hỗ trợ việc truy cập cho 25.000 sinh viên và gần 1.000 giảng viên, cán bộ viên chức của nhà trường.

- **Địa chỉ truy cập: http://csdl.hcmute.edu.vn/**
- Sử dụng email do nhà trường cấp để tạo tài khoản đăng ký:
  theo địa chỉ email của cán bộ **…@hcmute.edu.vn**
  theo địa chỉ email của sinh viên **…@student.hcmute.edu.vn**
- Hình thức truy cập: Theo IP đăng kí của nhà trường.
- ***Lưu ý:*** *Việc khai thác các CSDL được nhà xuất bản xác minh người dùng với nhiều lớp bảo mật qua dải địa chỉ IP và username/password, quý thầy cô và bạn đọc chỉ có thể khai thác các CSDL này trên hệ thống máy tính được kết nối mạng trong trường.*

Thông tin hỗ trợ kỹ thuật:

Trong quá trình sử dụng, quý thầy cô và bạn đọc cần hỗ trợ kỹ thuật xin vui lòng liên hệ:

Cô Trần Thị Ngọc Ý, Số ĐT 0919888975,
email: yttn@hcmute.edu.vn

Thư viện trân trọng thông báo đến Quý Thầy Cô và Các Bạn Sinh viên biết để sử dụng và khai thác các cơ sở dữ liệu trực tuyến đạt hiệu quả.

Trân trọng!

ĐỀ CƯƠNG CHI TIẾT MÔN HỌC

CÁC ỨNG DỤNG CỦA DỮ LIỆU LỚN - HỌC MÁY Ở QUY MÔ LỚN

BIG DATA APPLICATIONS - MACHINE LEARNING AT SCALE

Học phần cung cấp kiến thức nền tảng giúp người học hiểu được các ứng dụng của học máy trên dữ liệu lớn, cụ thể:

- Tổng quan về học máy và các ứng dụng của học máy.
- Tổng quan về học máy với quy mô lớn dùng Spark MLLib.
- Xây dựng, tinh chỉnh và áp dụng mô hình hồi quy.
- Xây dựng cây quyết định và làm mạnh nó dùng kỹ thuật ensemble learning.
- Hiểu và áp dụng thuật toán SVM.
- Hiểu và áp dụng thuật toán PCA kết hợp k-means.
- Hiểu và xây dựng hệ thống gợi ý.
- Hiểu và áp dụng các phương pháp xử lý văn bản và mô hình chủ đề LDA, Word2Vec.

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH KHOA CÔNG NGHỆ THÔNG TIN	**Ngành đào tạo: Kỹ thuật Dữ liệu** **Trình độ đào tạo: Đại học** **Chương trình đào tạo: Kỹ thuật Dữ liệu**

Đề cương chi tiết môn học

1. **Tên học phần:** CÁC ỨNG DỤNG CỦA DỮ LIỆU LỚN - HỌC MÁY Ở QUY MÔ LỚN

   Mã học phần: BDML434077
2. **Tên Tiếng Anh:** BIG DATA APPLICATIONS - MACHINE LEARNING AT SCALE
3. **Số tín chỉ:** 3 tín chỉ (2/1/6) (2 tín chỉ lý thuyết, 1 tín chỉ thực hành/thí nghiệm)

   Phân bố thời gian: 15 tuần (2 tiết lý thuyết + 1*2 tiết thực hành + 6 tiết tự học/tuần)

4. **Các giảng viên phụ trách học phần:**

   1/ GV phụ trách chính: ……..

2/ Danh sách giảng viên cùng GD: ……..

5. Điều kiện tham gia học tập học phần

Môn học tiên quyết: **Không**

Môn học trước: **Khai phá dữ liệu**

6. Mô tả học phần (Course Description)

Học phần cung cấp kiến thức nền tảng giúp người học hiểu được các ứng dụng của học máy trên dữ liệu lớn, cụ thể:

- Tổng quan về học máy và các ứng dụng của học máy.
- Tổng quan về học máy với quy mô lớn dùng Spark MLLib.
- Xây dựng, tinh chỉnh và áp dụng mô hình hồi quy.
- Xây dựng cây quyết định và làm mạnh nó dùng kỹ thuật ensemble learning.
- Hiểu và áp dụng thuật toán SVM.
- Hiểu và áp dụng thuật toán PCA kết hợp k-means.
- Hiểu và xây dựng hệ thống gợi ý.
- Hiểu và áp dụng các phương pháp xử lý văn bản và mô hình chủ đề LDA, Word2Vec.

7. Mục tiêu học phần (Course Goals)

Mục tiêu ***(Goals)***	**Mô tả** ***(Goal description)*** *(Học phần này trang bị cho sinh viên)*	**Chuẩn đầu ra CTĐT**	**Trình độ năng lực**
G1	Trình bày được các khái niệm cơ bản về dữ liệu lớn và học máy Nắm tổng quan về thư viện Spark MLLib	1.3 2.1, 2.3	2
G2	Khả năng cài đặt, áp dụng và đánh giá được các thuật toán học máy dùng Spark MLLib để phân tích dữ liệu lớn	2.1, 2.2, 2.3, 4.4	3
G3	Kỹ năng làm việc nhóm, và thuyết trình	2.4, 2.5, 3.1, 3.2	3

8. Chuẩn đầu ra của học phần

Chuẩn đầu ra HP		**Mô tả (Sau khi học xong môn học này, người học có thể)**	**Chuẩn đầu ra CTĐT**	**Trình độ năng lực**
G1	G1.1	Hiểu được ý nghĩa, tầm quan trọng và các thách thức của dữ liệu lớn.	1.3	2
	G1.2	Hiểu được các khái niệm cơ bản về học máy	1.3, 2.3	2
	G1.3	Hiểu được các thách thức khi học máy ở quy mô lớn	2.1	2
	G1.4	Hiểu được gì mà thư viện Spark MLLib hỗ trợ	2.3	2
G2	G2.1	Hiểu áp dụng được các thuật toán học máy	2.1, 2.3	3

		cho dữ liệu lớn		
	G2.2	Cài đặt và áp dụng được các thuật toán học máy cho dữ liệu lớn dùng Spark MLLib	2.1, 2.3, 4.4	3
	G2.3	Đánh giá được hiệu quả các thuật toán trên các tập dữ liệu thử nghiệm	2.1, 2.2	3
G3	G3.1	Làm việc hiệu quả trong một nhóm	3.1, 2.4, 2.5	3
	G3.2	Trình bày trước đám đông sử dụng phương tiện trình chiếu	3.2	3

9. Đạo đức khoa học:

Các bài tập ở nhà, bài kiểm tra phải được thực hiện từ chính bản thân sinh viên. Nếu bị phát hiện có sao chép thì xử lý các sinh viên có liên quan bằng hình thức đánh giá **0** (không) điểm quá trình và cuối kỳ.

10. Nội dung chi tiết môn học:

Tuần	Nội dung	Chuẩn đầu ra môn học	Trình độ năng lực	Phương pháp dạy học	Phương pháp đánh giá
1	***Chương 1: Tổng quan về học máy ở quy mô lớn*** *(4/0/8)*		2		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Các khái niệm cơ bản về dữ liệu lớn + Các khái niệm cơ bản về học máy o Học có giám sát (supervised learning) o Học không có giám sát (unsupervised learning) o Đánh giá mô hình ▪ Train/Validation/Test Sets ▪ Lựa chọn mô hình ▪ Bias vs. Variance o Các ứng dụng của học máy + Các vấn đề và thách thức khi học máy ở quy mô lớn + Tổng quan về thư viện Spark MLLib **PPGD chính** + Thuyết giảng + Thảo luận + Trình chiếu Powerpoint	G1.1 G1.2 G1.3 G1.4		- Đàm thoại - Thảo luận	Câu hỏi ngắn

	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc tài liệu tổng quan về học máy + **Tài liệu** o [1] Chương 1, 2 o [3] Chương 1 o [2] Chương 24, 25	G1.1 G1.2 G1.3 G1.4		Nêu và giải quyết vấn đề	Tự luận
2-3	***Chương 2: Hồi quy*** *(4/8/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Hồi quy tuyến tính o Giải thuật Gradient descent + Hồi quy logistic o Phân loại ở quy mô lớn o Giải thuật Gradient descent o Phân loại nhiều lớp (Multiclass Classification) + Regularization o Vấn đề quá khớp (overfitting) o Regularized Linear Regression o Regularized Logistic Regression + Ứng dụng của hồi quy **PPGD chính** + Thuyết giảng + Minh họa cách tính toán cho các mô hình + Trình chiếu Powerpoint	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc lại về mô hình boolean và mô hình không gian vector + Làm bài tập + Cài đặt, áp dụng hồi quy với Spark MLLib + **Tài liệu** o [1] Chương 3, 4 (mục 4.1 -> 4.3), 6 o [2] Chương 27	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận
4-5	***Chương 3: Cây quyết định và ensemble learning*** *(4/4/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Cây quyết định căn bản	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn

	+ Cây quyết định cho hồi quy + Cây quyết định cho phân loại + Boostrap và bagging + Gradient Boosted Decision Tree + Stochastic Boosting + Ứng dụng của cây quyết định và ensemble learning **PPGD chính** + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm				
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc tài liệu về cây quyết định và ensemble learning + Cài đặt, áp dụng cây quyết định và ensemble learning với Spark MLLib + **Tài liệu** o [1] Chương 8 o [2] Chương 26	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận
6-7	***Chương 4: Support vector machine - SVM*** *(4/4/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Cơ sở Toán cho SVM + Các khái niệm quan trọng của SVM o Mặt phẳng tách (separating hyperplane) o Tối ưu hóa biên của mặt phẳng tách (maximum margin separating hyperplane) o Biên mềm (soft margin) o Hàm nhân (kernel function) + Ứng dụng SVM **PPGD chính** + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn

	***B/* Các nội dung cần tự học ở nhà** *(8)* + Xem tài liệu về SVM + Áp dụng SVM dùng Spark MLLib + **Tài liệu** o [1] Chương 9	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận
8-9	***Chương 5: Gom cụm*** *(4/0/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Thu giảm số chiều o Phân tích thành phần chính (Principle Component Analysis - PCA) ▪ Thuật toán ▪ Chọn số thành phần chính ▪ Khôi phục dữ liệu từ các thành phần chính + Gom cụm k-means o Hàm mục tiêu o Thuật toán o Khởi tạo ngẫu nhiên o Chọn số cụm + Ứng dụng PCA và k-means + Đề xuất project **PPGD chính** + Thuyết giảng + Minh họa cách tính toán cho mô hình + Trình chiếu Powerpoint	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc tài liệu về PCA và k-means + Áp dụng PCA và k-means dùng Spark MLLib + **Tài liệu:** o [1] Chương 10 o [2] Chương 29	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận
10-11	***Chương 6: Hệ thống gợi ý*** *(4/4/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Giới thiệu hệ thống gợi ý	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn

	○ Hệ thống gợi ý phi cá nhân hóa ○ Hệ thống gợi ý dựa trên nội dung ○ Đánh giá hệ thống gợi ý + Hệ thống gợi ý lọc cộng tác (collaborative filtering) ○ User-user ○ Item-item ○ SVD ○ MF ○ ALS **PPGD chính** + Thuyết giảng + Minh họa cách tính toán cho mô hình + Trình chiếu Powerpoint				
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc tài liệu về hệ thống gợi ý + Áp dụng các mô hình gợi ý dùng Spark MLLib + **Tài liệu:** ○ [4] Chương 1->3, 5, 7 ○ [2] Chương 28	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận
12-13	***Chương 7: Học máy với dữ liệu văn bản*** *(4/4/8)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung GD lý thuyết** + Biểu diễn văn bản + Mô hình hóa chủ đề - topic modeling ○ LDA ○ Word2Vec + Ứng dụng mô hình chủ đề **PPGD chính** + Thuyết giảng + Minh họa kỹ thuật + Trình chiếu Powerpoint + Thảo luận nhóm	G3.1 G3.2 G3.3		- Đàm thoại - Thảo luận	Câu hỏi ngắn
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Đọc tài liệu về mô hình chủ đề + Áp dụng mô hình chủ đề dùng Spark MLLib để phân tích văn bản + **Tài liệu** ○ Các bài báo liên quan ○ [3] Chương 6	G3.1 G3.2 G3.3		Nêu và giải quyết vấn đề	Tự luận

14-15	***Chương 8: Các chủ đề chọn lọc*** *(8/0/16)*		3		
	***A/* Các nội dung và PPGD chính trên lớp** *(4)* **Nội dung chính trên lớp** + Các nhóm báo cáo các chủ đề đã chọn (từ tuần 8) **PPGD chính** + Trình chiếu Powerpoint + Thảo luận nhóm	G4.1 G4.2		- Đàm thoại - Thảo luận	Câu hỏi ngắn
	***B/* Các nội dung cần tự học ở nhà** *(8)* + Tìm kiếm tài liệu, đọc hiểu, viết báo cáo và trình bày một chủ đề đã chọn (giảng viên sẽ cung cấp các tài liệu cơ sở) + **Tài liệu** ○ [3] ○ [2] Chương 30, 31 ○ Các bài báo, sách liên quan về các chủ đề.	G4.1 G4.2			

11. Đánh giá kết quả học tập

- Thang điểm: **10**
- Kế hoạch kiểm tra như sau:

Hình thức KT	Nội dung	Thời điểm	Chuẩn đầu ra đánh giá	Trình độ năng lực	Phương pháp đánh giá	Công cụ KT	Tỉ lệ (%)
Bài tập - thực hành							**50**
BT#1	Bài tập về hồi quy	Tuần 3	G2.1 G3.1	3	Bài tập - thực hành	Câu hỏi	5
BT#2	Bài tập về phân loại và hồi quy dùng cây quyết định và ensemble learning	Tuần 5	G2.2 G3.1	3	Bài tập - thực hành	Câu hỏi	10
BT#3	Bài tập về phân loại và hồi quy dùng SVM	Tuần 7	G2.3 G3.1 G3.2	3	Bài tập - thực hành	Câu hỏi	10
BT#4	Bài tập về gom cụm dùng PCA và k-means	Tuần 9	G1.5 G3.1	3	Bài tập - thực hành	Câu hỏi	5
BT#5	Bài tập về hệ thống gợi ý	Tuần 11	G2.4 G3.1 G3.2	3	Bài tập - thực hành	Câu hỏi	10
BT#6	Bài tập về phân tích văn bản dùng mô hình chủ đề	Tuần 13	G2.4 G3.1 G3.2	3	Bài tập - thực hành	Câu hỏi	10

Thi cuối kỳ							**50**
	- Các nhóm (2-3 thành viên) trình bày một chủ đề chọn lọc về học máy ở quy mô lớn. - Mỗi thành viên sẽ được hỏi một số câu hỏi liên quan đến các kiến thức căn bản về học máy ở quy mô lớn.		G1.1->G1.6 G2.1->G2.6 G3.1->G3.2 G4.1->G4.2	3 3 3 3	Vấn đáp	Câu hỏi	50

12. Tài liệu học tập

- Sách, giáo trình chính:

[1] Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, *An Introduction to Statistical Learningwith Applications in R*, Springer, 2013. (http://www-bcf.usc.edu/~gareth/ISL/)

- Sách (TLTK) tham khảo:

[2] Bill Chambers and Matei Zaharia, *Spark - The Definitive Guide: Big Data Processing Made Simple*, O'Reilly Media, 2018.

[3] Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, *Advanced Analytics with Spark: Patterns for Learning from Data at Scale*, O'Reilly Media, 2015.

[4] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich, *Recommender Systems: An Introduction*, Cambridge University Press, 2011.

13. Ngày phê duyệt lần đầu:

14. Cấp phê duyệt:

Trưởng khoa **Trưởng BM** **Nhóm biên soạn**

15. Tiến trình cập nhật ĐCCT

Lần 1: Nội Dung Cập nhật ĐCCT (người cập nhật ký và ghi rõ họ tên)

lần 1: ngày tháng năm

Tổ trưởng Bộ môn:

GIỚI THIỆU CÁC CHƯƠNG TRÌNH ĐÀO TẠO

1. Ngành Công nghệ Thông tin (7480201D, 7480201C, 7480201A)

Sinh viên có kiến thức chuyên môn vững chắc về lĩnh vực Công nghệ thông tin (Công nghệ phần mềm, Mạng máy tính, Hệ thống thông tin); Sinh viên được chú trọng học tập cả về lý thuyết lẫn kỹ năng thực hành nhằm đảm bảo Sinh viên ra trường có thể tiếp cận và làm việc ngay tại các công ty.

Ra trường, Sinh viên có thể làm việc tại: Các công ty phần mềm, các công ty cung cấp giải pháp mạng, các công ty cung cấp giải pháp hệ thống thông tin trong nước cũng như quốc tế,… Có thể tham gia nghiên cứu và giảng dạy tại các Viện nghiên cứu, Trường Đại học, Cao đẳng, Trung cấp chuyên nghiệp, cũng như các Trường phổ thông; có thể làm việc tại các công ty ngoài lĩnh vực CNTT nhưng có áp dụng triển khai CNTT trong hoạt động, quản lý, sản xuất.

Hiện nay nhu cầu về kỹ sư phần mềm tại Việt Nam là rất lớn bởi sự chuyển hướng đầu tư của các công ty nước ngoài vào Việt Nam. Ngoài ra, SV sau khi hoàn thành chương trình đại học, có khả năng học tiếp lên các bậc học cao hơn (thạc sỹ, tiến sỹ).

Được đào tạo tại khoa Công nghệ Thông tin và khoa Đào tạo Chất lượng cao.

2. Ngành Kỹ thuật dữ liệu (7480203D)

Ngành Kỹ thuật dữ liệu trang bị cho sinh viên cho sinh viên kiến thức cơ sở ngành về công nghệ thông tin, và kiến thức chuyên sâu về phân tích, thiết kế, xây dựng các mô hình vật lý và mô hình lô-gic cho dữ liệu, xây dựng hạ tầng dữ liệu (data infrastructure, data sets) và các công cụ phục vụ việc thiết kế, dự đoán các sản phẩm tương lai.

Các kiến thức được trang bị cho sinh viên như: Cấu trúc dữ liệu và giải thuật, lập trình hướng đối tượng, lập trình web, trí tuệ nhân tạo, Khai phá dữ liệu, học máy, mô hình rủi ro tài chính, v.v…

Sinh viên tốt nghiệp có thể làm việc trong các công ty Công nghệ thông tin, cũng như bộ phận công nghệ thông ty của các công ty ở các lĩnh vực khác. Các vị trí có thể đảm nhận trong các tổ chức gồm:

- Data Engineer (Kỹ sư dữ liệu).
- Big data Engineer (Kỹ sư quản trị dữ liệu lớn).
- Lập trình viên, chuyên viên phát triển phần mềm.
- Chuyên viên quản trị cơ sở dữ liệu, quản trị web.
- Trưởng dự án phát triển hệ thống thông tin.

- Cán bộ nghiên cứu và ứng dụng Công nghệ thông tin ở các viện, trung tâm nghiên cứu và các trường đại học, cao đẳng.
- Giảng viên Công nghệ thông tin ở các các trường đại học, cao đẳng, trung học phổ thông, trung học chuyên nghiệp và dạy nghề.

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH

www.hcmute.edu.vn

ISBN: 978-604-73-2175-9

9 786047 321759