

Trường Đại học Bách Khoa -
ĐHQG Tp.HCM
Khoa: Phòng Đào tạo Sau đại
học
Khoa/Bộ môn quản lý MH:
Phòng Đào tạo Sau đại học

Tp.HCM, ngày tháng
năm

Đề cương môn học Sau đại học

CHUYÊN ĐỀ NGHIÊN CỨU: PHÂN TÍCH DỮ LIỆU (RESEARCH TOPIC: DATA ANALYTICS)

Mã số MH: GK5938

Số tín chỉ:		Tc (LT.BT&TH.Tự Học): 3					ECTS: 6					
Số tiết	-Tổng:	75	LT:	30	BT:	0	TH:	0	ĐA:		BTL/TL:	45
Đánh giá:	Thuyết trình các chủ đề:				20%	Theo nhóm						
	Tiểu luận:				30%	Theo nhóm						
	Thi cuối kỳ:				50%	Trắc nghiệm + tự luận						
- Môn tiên quyết:												
- Môn học trước:												
- Môn song hành:												
- CTĐT ngành (Mã ngành):		Kỹ Thuật Hóa Học (8520301), Công Nghệ Sinh Học (8420201),...										
- Ghi chú khác:												

1. Mục tiêu môn học:

Môn học này nhằm đào tạo sinh viên kiến thức và kỹ năng về phân tích dữ liệu (PTDL - data analytics) như: quy trình và nguyên lý phân tích dữ liệu, các kỹ thuật thu thập, tiền xử lý dữ liệu, các kỹ thuật phân tích dữ liệu (data analysis), ứng dụng của phân tích dữ liệu trong khoa học - kỹ thuật - quản lý.

Môn học được áp dụng cho sinh viên sau đại học của tất cả các ngành khi có mong muốn nâng cao kiến thức, kỹ năng, định hướng nghiên cứu và áp dụng phân tích dữ liệu, khoa học dữ liệu vào thực tiễn.

Aims:

This course provides students knowledge and skills in data analytics including processes and principles in data analytics, data gathering and pre-processing techniques, data analysis techniques, their applications in science, technology and management.

This module is applied to graduate students in any major who want to improve knowledge, skill, research direction and applications of data analytics and data science to the practices.

2. Nội dung tóm tắt môn học:

Môn học này trình bày các nội dung chính sau:

- Các nguyên lý và quy trình trong phân tích dữ liệu
- Các kỹ thuật thu thập và tiền xử lý dữ liệu
- Phân tích dữ liệu dựa vào thống kê
- Giới thiệu về Phân tích dữ liệu dựa vào các kỹ thuật khai phá dữ liệu: dự báo (prediction), phân lớp (classification), gom cụm (clustering), khai phá luật kết hợp (association rule mining)
- Giới thiệu về Kho dữ liệu
- Giới thiệu về Hệ cơ sở dữ liệu và các mô hình dữ liệu
- Nghiên cứu, thảo luận về các vấn đề hiện đại trong phân tích dữ liệu, ứng dụng thực tiễn và các hướng nghiên cứu chuyên sâu (thông qua tiểu luận)

Course outline:

In details, following topics will be presented in the module:

Principles and processes in data analytics

Data collection and pre-processing techniques

Statistics and data analysis

Data analysis based on data mining technique: prediction, classification, clustering, association rule mining

Introduction to data organization and models: data warehouse, database systems

Investigate and discuss on modern issues in data analytics, its applications and research directions.

3. Tài liệu học tập:

Giáo trình/Textbook

- [1] [Anil Maheshwari](#) “Data Analytics: Made accessible”, Public book, 2015.
- [2] [EMC](#) Education Services “Data Science & Big Data Analytics”, John Wiley and Sons Inc, 2015.
- [3] Trần Minh Quang, "Khai Phá Dữ Liệu và Kỹ Thuật Phân Lớp", NXB Đại Học Quốc Gia TP. HCM, 2020.
- [4] Christian Heumann, Michael Schomaker “Introduction to Statistics and Data Analysis”, Springer, 2016.

Sách tham khảo/References

- [5] Foster Provost, Tom Fawcett “Data Science for Business”, O'Reilly, 2013.
- [6] Fabio Nelli, “Python Data Analytics”, Apress, 2015.
- [7] Thomas Mailund, “Beginning Data Science in R”, Apress, 2017.

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CDR Chương trình (CĐRCT)	
			Ứng dụng	Nghiên cứu

Learning outcomes:

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO	
			Coursework	Research

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình ứng dụng:

	Chuẩn đầu ra của chương trình (CĐRCT)										
Chuẩn đầu ra môn học (CĐRMH)	a	b	c	d	e	f	g	h	i	j	k

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình nghiên cứu:

	Chuẩn đầu ra của chương trình (CĐRCT)										
Chuẩn đầu ra môn học (CĐRMH)	a	b	c	d	e	f	g	h	i	j	k

5. Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Đọc hiểu, phân tích các kỹ thuật về phân tích dữ liệu trong giáo trình (text book), tài liệu tham khảo và trên Internet

Khảo sát các ứng dụng thực tế của phân tích dữ liệu

Thực hành lập trình hoặc sử dụng tốt các công cụ phân tích dữ liệu (ví dụ Python, R, Weka, Orange...)

Mở rộng các vấn đề nghiên cứu trong phân tích dữ liệu

Về thực hiện báo cáo tiểu luận: nhóm 1-2 người, bắt đầu thực hiện từ tuần 1 đến tuần 15, nộp đề cương của báo cáo tiểu luận vào tuần 4 (không bắt buộc), nộp báo cáo tiểu luận vào tuần 15 và xung phong trình bày báo cáo tiểu luận vào tuần 15.

Cách đánh giá: 20%*điểm kiểm tra trên lớp bao gồm trình bày các chủ đề + 30%*điểm báo cáo tiểu luận + 50%*điểm thi cuối kỳ từ 5.5 trở lên (≥ 5.5) mới tính là đạt cả môn học.

Learning strategies & Assessment Scheme:

Comprehensive reading and analyzing data analytics method and mechanisms from text book and for the Internet

Investigate real-world applications where data analytics can be applied

Practice with programming and tools (e.g., R, Python, Weka, Orange,...) for data analytics

Discuss on research issues and development trends in this field

Students need to read textbooks and finish all assignments.

As for seminar report, each group includes one-two students, starts working from week 1 to week 15, optionally submits the description of a seminar report in week 4, submits the final version of a seminar report in week 15, and optionally presents a seminar report in week 15.

Assessment scheme:

20% * in-class exam score including presentations + 30% * Mini project score + 50% * Final exam score ≥ 5.5 in order to pass this course.

6. Nội dung chi tiết:

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Chương 1: Tổng quan về phân tích dữ liệu	Chương 1: Tổng quan về phân tích dữ liệu 1.1. Vai trò của phân tích dữ liệu 1.2. Phân tích dữ liệu và hỗ trợ ra quyết định, BI. Quá trình phân tích dữ liệu - Phân biệt quá trình phân tích dữ liệu (data analytics) và các kỹ thuật phân tích dữ liệu (analysis)- Các khái niệm quan trọng trong quá trình phân tích dữ liệu như: dữ liệu (data), hệ CSDL (database), kho dữ liệu (data warehouse), khai phá dữ liệu (data mining), biểu diễn dữ liệu (data visualizatio),...		[1, 2, 4]
2	Chương 2: Tiền xử lý dữ liệu bằng thống kê mô tả (descriptive statistics)	Chương 2: Tiền xử lý dữ liệu bằng thống kê mô tả (descriptive statistics) 2.1. Tổng quan về giai đoạn tiền xử lý dữ liệu 2.2. Tóm tắt hoá 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 2.9. Biểu diễn dữ liệu		[1, 2, 3, 5]
3	Chương 3: Phân tích dữ liệu bằng các kỹ thuật thống kê qui nạp (inductive statistics)	Chương 3: Phân tích dữ liệu bằng các kỹ thuật thống kê qui nạp (inductive statistics) 3.1. Giới thiệu về thống kê quy nạp và suy luận (inference) 3.2. Đặc tính của công cụ ước lượng điểm 3.3. Ước lượng khoảng 3.4. Ước lượng khoảng 3.5. Xác định kích thước mẫu 3.6. Các vấn đề thảo luận mở rộng		[3, 4]
4	Chương 4: Giới thiệu Khai phá dữ liệu	Chương 4: Giới thiệu Khai phá dữ liệu 4.1. Tổng quan về quá trình khám phá tri thức 4.2. Các khái niệm 4.3. Ý nghĩa và vai trò của khai phá dữ liệu 4.4. Ứng dụng của khai phá dữ liệu		[1, 3, 5]
5	Chương 5: Kỹ thuật dự báo dựa vào hồi qui	Chương 5: Kỹ thuật dự báo dựa vào hồi qui 5.1. Tổng quan về hồi qui 5.2. Hồi qui tuyến tính 5.3. Hồi qui phi tuyến 5.4. Ứng dụng 5.5. Các vấn đề với hồi qui		[1, 3]
6, 7	Chương 6: Phân lớp	Chương 6: Phân lớp 6.1. Tổng quan về phân lớp 6.2. Phân lớp với cây quyết định 6.3. Phân lớp bằng phương pháp Bayesian 6.4. Phân lớp với mạng Neural (đọc thêm) 6.5. Các phương pháp phân lớp khác		[1-3]

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
8	Chương 7: Giới thiệu về Gom cụm dữ liệu	Chương 7: Giới thiệu về Gom cụm dữ liệu 7.1. Tổng quan về gom cụm dữ liệu 7.2. Phân loại các kỹ thuật gom cụm (bằng phân hoạch, bằng phân cấp, dựa trên mật độ, dựa trên mô hình) 7.3. Giải thuật K-means 7.4. Các kỹ thuật gom cụm khác (sinh viên đọc thêm)		[1-3]
9	Chương 8: Giới thiệu về khai phá luật kết hợp	Chương 8: Giới thiệu về khai phá luật kết hợp 8.1. Tổng quan về luật kết hợp 8.2. Biểu diễn luật kết hợp 8.3. Khám phá các mẫu thường xuyên 8.4. Giải thuật Apriori 8.5. Các giải thuật khác (sinh viên đọc thêm)		[1-3]
10	Chương 9: Giới thiệu về Hệ CSDL và Kho dữ liệu	Chương 9: Giới thiệu về Hệ CSDL và Kho dữ liệu 9.1. Vai trò của Hệ CSDL và Các khái niệm 9.2. Vai trò và Các khái niệm về Kho dữ liệu 9.3. Giới thiệu về các mô hình CSDL- Mô hình CSDL quan hệ- Mô hình ER/EER (để thiết kế ý niệm) 9.4. Các cách tiếp cận thiết kế Kho dữ liệu: Mô hình hình sao		[1-3]
11 - 15	Tiểu luận	Tiểu luận - Làm việc theo nhóm xây giải quyết các chủ đề liên quan về phân tích dữ liệu- Trình bày trước lớp, có phản biện		[1-8]

7. Giảng viên tham gia giảng dạy:

CBGD
chính:

PGS.TS
Trần
Mình
Quang

CBGD
tham
gia:

PGS.TS
Võ Thị
Ngọc
Châu
TS.
Phan
Trọng
Nhân
PGS.TS
Lê
Hong
Trang

**XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA**

*Tp. Hồ Chí
Minh, ngày
..... tháng
..... năm*

.....
**GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG**

**PGS.TS
Trần
Minh
Quang**