

HỌC PHẦN
KHAI PHÁ DỮ LIỆU

(Ban hành kèm theo Quyết định số .../... ngày 11 tháng 5 năm 2023
của Hiệu trưởng Trường Đại học Sài Gòn)

1. Thông tin tổng quát về học phần:

- Tên học phần (tiếng Việt): Khai phá dữ liệu (tiếng Anh): Data Mining	
- Mã số học phần: 858016	
- Thuộc khối kiến thức/ kỹ năng: <input type="checkbox"/> Kiến thức giáo dục đại cương <input checked="" type="checkbox"/> Kiến thức ngành <input type="checkbox"/> Kiến thức cơ sở của ngành <input type="checkbox"/> Kiến thức chuyên ngành (nếu có)	
- Số tín chỉ:	3
+ Số tiết lý thuyết:	30
+ Số tiết thảo luận/bài tập:	0
+ Số tiết thực hành:	30
+ Số tiết hoạt động nhóm:	0
+ Số tiết tự học:	90
- Học phần tiên quyết:	858015 Máy học
- Học phần song hành (nếu có):	Không

2. Mô tả học phần (Vị trí, vai trò của học phần đối với chương trình đào tạo và khái quát những nội dung chính)

Học phần giảng dạy các kiến thức nâng cao về máy học, bao gồm:

- Các khái niệm và ứng dụng của khai phá dữ liệu.
- Các thuật toán phân lớp và gom nhóm dữ liệu.
- Các thuật toán khai phá luật kết hợp.
- Triển khai hệ thống khai phá dữ liệu.

3. Mục tiêu học phần

Học phần giúp sinh viên đạt được những mục tiêu sau:

Về kiến thức: Sinh viên được trang bị các kiến thức chuyên môn cũng như các phương pháp và kỹ thuật thực hành để ứng dụng các kiến thức đã học vào việc giải quyết các bài toán trong thực tiễn.

Về kỹ năng: Sinh viên được trau dồi các kỹ năng tìm kiếm và đọc tài liệu, kỹ năng làm việc nhóm, kỹ năng thuyết trình, tư duy phản biện.

Về thái độ: Sinh viên có tinh thần học tập chuyên cần, nghiêm túc; tác phong chuyên nghiệp.

4. Chuẩn đầu ra học phần

Ký hiệu chuẩn đầu ra (1)	Mô tả chuẩn đầu ra (2)	Trình độ năng lực (3)
G1	Hiểu (giải thích, mô tả, nhận xét, ...) các kiến thức nền tảng của học phần	3.0

G2	Áp dụng (vận dụng, chỉ ra, minh họa, ...) các kiến thức đã học để giải quyết các bài toán thực tiễn	3.5
G3	Phân tích (phân biệt, phân tích, khảo sát, ...) đặc điểm của các phương pháp, kỹ thuật và khả năng ứng dụng chúng trong các bài toán thực tiễn	4.0
G4	Đánh giá (đánh giá, so sánh, liên hệ, nhận định, ...) hiệu quả của các phương pháp, kỹ thuật	4.5
G5	Đề xuất và xây dựng giải pháp cho các bài toán mới trong thực tiễn	5.0

5. Nội dung chi tiết học phần

CHƯƠNG 1. GIỚI THIỆU KHAI PHÁ DỮ LIỆU

- 1.1. Lịch sử khai phá dữ liệu.
- 1.2. Quy trình khai phá dữ liệu.
- 1.3. Các ứng dụng của khai phá dữ liệu.

CHƯƠNG 2. KHAI PHÁ DỮ LIỆU

- 2.1. Các phương pháp khai phá dữ liệu.
- 2.2. Xử lý dữ liệu thô và lựa chọn thuộc tính.
- 2.3. Chuẩn hóa dữ liệu.

CHƯƠNG 3. PHÂN LỚP VÀ GOM NHÓM DỮ LIỆU

- 3.1. Giới thiệu bài toán phân lớp và gom nhóm dữ liệu.
- 3.2. Các thuật toán phân lớp: Naive Bayes, Decision Trees, Random Forest, Support Vector Machines.
- 3.3. Các thuật toán gom nhóm: K-Means, Hierarchical Clustering.

CHƯƠNG 4. KHAI PHÁ LUẬT KẾT HỢP

- 4.1. Các khái niệm và định nghĩa.
- 4.2. Các thuật toán: Apriori, FP-Growth.

CHƯƠNG 5. TRIỂN KHAI HỆ THỐNG KHAI PHÁ DỮ LIỆU

- 5.1. Tiền xử lý dữ liệu
- 5.2. Mô hình hóa
- 5.3. Kiểm thử mô hình và đánh giá

6. Học liệu

6.1. Tài liệu bắt buộc (không quá 3 tài liệu)

- [1] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. (2006). Introduction to Data Mining. Pearson Education.
- [2] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques (3rd ed.). San Francisco, CA: Morgan Kaufmann Publishers.

6.2. Tài liệu tham khảo

- [1] Ian H. Witten, Eibe Frank, and Mark A. Hall. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers.
- [2] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Mining of Massive Datasets. Cambridge University Press.

7. Hướng dẫn tổ chức dạy học

Tuần/ Buổi học	Nội dung dạy học	Hình thức tổ chức dạy học	Yêu cầu đối với sinh viên	CĐR môn học
1	CHƯƠNG 1. GIỚI THIỆU KHAI PHÁ DỮ LIỆU 1.1. Lịch sử khai phá dữ liệu. 1.2. Quy trình khai phá dữ liệu. 1.3. Các ứng dụng của khai phá dữ liệu.	Giảng viên thuyết giảng, hướng dẫn bài tập; sinh viên làm bài tập, thuyết trình, làm việc nhóm	Sinh viên đọc trước tài liệu và xem lại bài sau mỗi buổi học, làm bài tập tại lớp và ở nhà, làm bài thuyết trình/đồ án...	G1 G2 G3 G4
2-3	CHƯƠNG 2. KHAI PHÁ DỮ LIỆU 2.1. Các phương pháp khai phá dữ liệu. 2.2. Xử lý dữ liệu thô và lựa chọn thuộc tính. 2.3. Chuẩn hóa dữ liệu.			G1 G2 G3 G4
4-6	CHƯƠNG 3. PHÂN LỚP VÀ GOM NHÓM DỮ LIỆU 3.1. Giới thiệu bài toán phân lớp và gom nhóm dữ liệu. 3.2. Các thuật toán phân lớp: Naive Bayes, Decision Trees, Random Forest, Support Vector Machines. 3.3. Các thuật toán gom nhóm: K-Means, Hierarchical Clustering.			G1 G2 G3 G4
7-9	CHƯƠNG 4. KHAI PHÁ LUẬT KẾT HỢP 4.1. Các khái niệm và định nghĩa. 4.2. Các thuật toán: Apriori, FP-Growth.			G1 G2 G3 G4 G5
10-13	CHƯƠNG 5. TRIỂN KHAI HỆ THỐNG KHAI PHÁ DỮ LIỆU 5.1. Tiền xử lý dữ liệu 5.2. Mô hình hóa 5.3. Kiểm thử mô hình và đánh giá			G1 G2 G3 G4 G5
14-15	BÁO CÁO TIỂU LUẬN	Giảng viên nhận xét và đánh giá tiểu luận của sinh viên	Sinh viên báo cáo tiểu luận	G1 G2 G3 G4

8. Quy định đối với môn học và yêu cầu của giảng viên (những yêu cầu khác đối với môn học (nếu có), phù hợp với các quy chế đào tạo hiện hành)

- Sinh viên có mặt tối thiểu 80% số buổi học trên lớp.
- Sinh viên tuân thủ các yêu cầu của giảng viên trong mỗi buổi học.

9. Phương pháp đánh giá học phần

9.1. Thang điểm và cách tính điểm đánh giá

Điểm đánh giá bộ phận và điểm học phần được tính theo thang điểm 10 (từ 0 đến 10), làm tròn đến một chữ số thập phân

9.2. Đánh giá bộ phận

Bộ phận được đánh giá	Điểm đánh giá bộ phận	Trọng số	Hình thức đánh giá
1. Đánh giá quá trình	Điểm quá trình	0.5 (1.1 + 1.2)	
1.1. Ý thức học tập	Điểm chuyên cần, thái độ học tập, ...	0.1	Điểm danh; các hoạt động học tập tại lớp
1.2. Hồ sơ học tập	- Điểm bài tập ở nhà và trên lớp, bài tập lớn, ... - Điểm thuyết trình, thực hành, thảo luận, làm việc nhóm, kiểm tra giữa kỳ, ...	0.2 0.2	
2. Đánh giá cuối kỳ	Điểm thi kết thúc học phần	0.5 (≥ 0.5)	Tiểu luận

9.3. Điểm học phần

Điểm học phần là điểm trung bình chung theo trọng số tương ứng của Điểm quá trình (Điểm đánh giá quá trình) và Điểm thi kết thúc học phần (Điểm đánh giá cuối kỳ).

10. Phụ trách học phần

- Khoa/bộ môn phụ trách: Khoa Công nghệ thông tin, Bộ môn Hệ thống thông tin
- Địa chỉ/email: 273 An Dương Vương, Phường 3, Quận 5, Tp. Hồ Chí Minh/Email: vpkcntt@sgu.edu.vn.

TP. Hồ Chí Minh, ngày 11 tháng 05 năm 2023

TRƯỞNG KHOA

TRƯỞNG BỘ MÔN

GIẢNG VIÊN BIÊN SOẠN

PGS.TS. Phạm Thế Bảo

PGS. TS. Nguyễn Tuấn Đăng

PGS. TS. Nguyễn Tuấn Đăng

DUYỆT

BAN GIÁM HIỆU



PGS.TS. Lê Minh Triết

PHỤ LỤC

GIẢI THÍCH MỘT SỐ KÝ HIỆU

- (1) Ký hiệu CĐR bằng các ký hiệu G từ 1,2,...;
- (2) Mô tả CĐR theo thang Bloom;
- (3) Thang trình độ năng lực:

Trình độ năng lực	Mô tả
0.0 -> 2.0	Nhớ (trình bày, định nghĩa, liệt kê,...)
2.0 -> 3.0	Hiểu (giải thích, mô tả, nhận xét,...)
3.0 -> 3.5	Áp dụng (vận dụng, chỉ ra, minh họa,...)
3.5 -> 4.0	Phân tích (phân biệt, phân tích, khảo sát,...)
4.0 -> 4.5	Đánh giá (đánh giá, so sánh, liên hệ, nhận định,...)
4.5 -> 5.0	Sáng tạo (thiết kế, đề xuất, tổ chức,...)