



ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

**ĐỀ CƯƠNG MÔN HỌC**  
**IE403 – KHAI THÁC DỮ LIỆU TRUYỀN THÔNG XÃ HỘI**

**1. THÔNG TIN CHUNG (General information)**

Tên môn học (tiếng Việt):	Khai thác dữ liệu truyền thông xã hội
Tên môn học (tiếng Anh):	Social Media Mining
Mã môn học:	IE403
Thuộc khối kiến thức:	Đại cương <input type="checkbox"/> ; Cơ sở nhóm ngành <input type="checkbox"/> ; Cơ sở ngành <input type="checkbox"/> ; Chuyên ngành <input type="checkbox"/> ; Tốt nghiệp <input checked="" type="checkbox"/>
Khoa phụ trách:	Khoa học và Kỹ thuật Thông tin
Giảng viên biên soạn:	ThS. Nguyễn Văn Kiệt. Email: <a href="mailto:kietnv@uit.edu.vn">kietnv@uit.edu.vn</a>
Số tín chỉ:	
Lý thuyết:	3
Thực hành:	0
Tự học:	6
Môn học tiên quyết:	Không.
Môn học trước:	Không.

**2. MÔ TẢ MÔN HỌC (Course description)**

Môn học này cung cấp kiến thức về truyền thông xã hội và các kỹ thuật khai thác thông tin từ truyền thông xã hội để phục vụ cho lợi ích của cơ quan, doanh nghiệp. Môn học cũng trang bị cho sinh viên một số kỹ thuật để xử lý được những dạng dữ liệu truyền thông xã hội khác nhau.

**3. MỤC TIÊU MÔN HỌC (Course goals)**

Sau khi hoàn thành môn học này, sinh viên có thể:

Bảng 1. Mục tiêu môn học.

Ký hiệu	Mục tiêu môn học	Chuẩn đầu ra trong CTĐT
G1	Nắm vững một số kiến thức nền tảng về xử lý và khai thác dữ liệu truyền thông xã hội.	LO 2
G2	Hiểu, phân tích, chọn phương pháp, thực nghiệm và đánh giá một nghiên cứu đơn giản thông qua bài toán hệ khuyến nghị, xử lý văn bản truyền thông và rút trích thông tin.	LO 2, LO3, LO4, LO 10

G3	Có kỹ năng làm việc và tổ chức nhóm	LO 7
----	-------------------------------------	------

#### 4. CHUẨN ĐẦU RA MÔN HỌC (Course learning outcomes)

Bảng 2. Chuẩn đầu ra môn học.

CĐRMH	Mô tả CĐRMH (Mục tiêu cụ thể)	Mức độ giảng dạy
G1.1	Nắm vững một số kiến thức nền tảng về xử lý và phân tích dữ liệu truyền thông xã hội.	IT
G1.1	Vận dụng kiến thức nền tảng về khai phá dữ liệu truyền thông xã hội vào quy trình xây dựng một hệ thống đi từ nghiên cứu cho đến ứng dụng.	IT
G2.1	Hiểu, phân tích, chọn phương pháp, thực nghiệm và đánh giá một nghiên cứu đơn giản thông qua bài toán hệ khuyến nghị, xử lý văn bản truyền thông và rút trích thông tin.	ITU
G2.2	Hiểu được các chủ thể liên quan đến doanh nghiệp và vai trò của thông tin truyền thông xã hội đối với doanh nghiệp. Có thể ứng dụng những kỹ thuật đã học để xây dựng kế hoạch thực hiện khai thác truyền thông xã hội cho doanh nghiệp.	IT
G5	Có kỹ năng làm việc và tổ chức nhóm	ITU

#### 5. NỘI DUNG MÔN HỌC, KẾ HOẠCH GIẢNG DẠY (Course content, lesson plan)

##### a. Lý thuyết

Bảng 3. Nội dung và kế hoạch giảng dạy lý thuyết của môn học.

Buổi học (45 tiết)	Nội dung	CĐRMH	Hoạt động dạy và học	Thành phần đánh giá
Buổi 1 (4 tiết)	<b>Chương 1: Giới thiệu về khai phá dữ liệu truyền thông xã hội</b> 1. Khai phá dữ liệu truyền thông xã hội là gì? 2. Những thách thức mới trong khai phá dữ liệu truyền thông xã hội. 3. Bài tập	G1	Dạy: Thuyết giảng Học ở nhà: ôn lại bài học và làm bài tập	A2, A4
Buổi 2 (4 tiết)	<b>Chương 2: Những dạng dữ liệu truyền thông xã hội</b> 1. Các dạng dữ liệu truyền thông xã hội. 2. Giá trị của dữ liệu truyền thông xã hội.	G1	Dạy: Thuyết giảng Học ở nhà: ôn lại bài học và làm bài tập	A1, A2, A4

	<p>3. Những ứng dụng dữ liệu truyền thông xã hội.</p> <p>4. Bài tập.</p> <p>5. <i>Đăng ký đề tài Seminar 1 &amp; 2.</i></p>			
Buổi 3 (4 tiết)	<p><b>Chương 3: Các mô hình khai phá dữ liệu</b></p> <p>1. Tại sao phải khai phá dữ liệu?</p> <p>2. Các phương pháp học máy.</p> <p>3. Học giám sát.</p> <p>4. Học không giám sát.</p> <p>5. Học sâu.</p> <p>6. Bài tập.</p>	G1	<p>Dạy: Thuyết giảng</p> <p>Học ở nhà: ôn lại bài học và làm bài tập</p>	A1, A2, A4
Buổi 4 (4 tiết)	<p><b>Chương 4: Quy trình nghiên cứu khai phá dữ liệu truyền thông xã hội</b></p> <p>1. Tạo bộ dữ liệu chuẩn như thế nào?</p> <p>2. Mô hình đề xuất.</p> <p>3. Thực nghiệm và đánh giá.</p> <p>4. Ứng dụng mô hình kết quả vào các ứng dụng thực tế.</p> <p>5. Bài tập.</p>	G1, G2	<p>Dạy: Thuyết giảng</p> <p>Học ở nhà: ôn lại bài học và làm bài tập</p>	A1, A2, A4
Buổi 5, 6 (8 tiết)	<p><b>Chương 5: Chuyên đề về hệ khuyến nghị dựa trên khai phá dữ liệu kinh doanh trên phương tiện truyền thông xã hội</b></p> <p>1. Tại sao phải xây dựng hệ khuyến nghị?</p> <p>2. Xây dựng dữ liệu.</p> <p>3. Các mô hình khuyến nghị.</p> <p>4. Thực nghiệm và đánh giá những khuyến nghị.</p> <p>5. Ứng dụng.</p> <p>6. <i>Báo cáo tiến độ đề tài Seminar 1.</i></p>	G3, G4, G5	<p>Dạy: Thuyết giảng</p> <p>Học ở nhà: ôn lại bài học và làm bài tập</p>	A1, A2, A4

Buổi 7, 8 (8 tiết)	<b>Chương 6: Chuyên đề về khai phá dữ liệu văn bản truyền thông xã hội</b> 1. Tại sao phải khai phá dữ liệu văn bản truyền thông xã hội. 2. Xây dựng dữ liệu. 3. Các mô hình khai phá văn bản truyền thông xã hội. 4. Thực nghiệm và đánh giá 5. Ứng dụng 6. Báo cáo tiến độ đề tài Seminar 2.	G3, G4, G5	Dạy: Thuyết giảng Học ở nhà: ôn lại bài học và làm bài tập	A1, A4
Buổi 9, 10 (8 tiết)	<b>Chương 7: Chuyên đề về rút trích thông tin tự động</b> 1. Tại sao phải rút trích thông tin. 2. Xây dựng dữ liệu. 3. Các mô hình rút trích thông tin 4. Thực nghiệm và đánh giá 5. Ứng dụng 6. Báo cáo tiến độ đề tài Seminar 3.	G3, G4, G5	Dạy: Thuyết giảng Học ở nhà: ôn lại bài học và làm bài tập	A1, A4
Buổi 11 (5 tiết)	<b>Báo cáo các chuyên đề</b>	G3, G4, G5	Từng nhóm sinh viên báo cáo.  Giảng viên: nhận xét và đánh giá.	A1, A4

**b. Thực hành: Không có**

**6. ĐÁNH GIÁ MÔN HỌC (Course assessment)**

Bảng 5. Các thành phần đánh giá của môn học.

Thành phần đánh giá	CĐRMH	Tỷ lệ (%)
A1. Quá trình	G1, G2, G3, G4, G5	50%
A2. Giữa kỳ		0%
A3. Thực hành		0%
A4. Cuối kỳ	G3, G4, G5	50%

**a. Ribric của thành phần đánh giá A1**

Hình thức: đồ án (project) + sự hiện diện.

Bảng 6. Định lượng và định tính của thành phần đánh giá A1.

<b>Tiêu chí</b>	<b>10</b>	<b>8</b>	<b>6</b>	<b>4</b>	<b>0</b>
Tinh thần làm việc nhóm (Dựa trên đánh giá của các thành viên trong nhóm)	Hoàn thành nhiệm vụ ~ $\geq 80\%$	Hoàn thành nhiệm vụ ~ [60 – 80%)	Hoàn thành nhiệm vụ ~ [40 – 60%)	Hoàn thành nhiệm vụ ~ [20 - 40%)	Hoàn thành nhiệm vụ ~ [0 - 20%)
Báo cáo viết	Hình thức trình bày: không vi phạm các qui tắc trình bày đã được hướng dẫn tại lớp.  Nội dung: đầy đủ các thành phần theo thứ tự bắt buộc.	Hình thức trình bày: vi phạm tối đa 01 qui tắc trình bày đã được hướng dẫn tại lớp.  Nội dung: đầy đủ các thành phần nhưng không theo thứ tự.	Hình thức trình bày: vi phạm tối đa 02 qui tắc trình bày đã được hướng dẫn tại lớp.  Nội dung: thiếu tối đa 01 thành phần bắt buộc.	Hình thức trình bày: vi phạm hơn 02 qui tắc trình bày đã được hướng dẫn tại lớp.  Nội dung: thiếu tối đa 02 thành phần bắt buộc.	Không làm
Thuyết trình	Nội dung: có đầy đủ các thành phần báo cáo theo đúng thứ tự: ngữ cảnh, mục tiêu, phương pháp, thực	Nội dung: có đầy đủ các thành phần báo cáo: ngữ cảnh, mục tiêu, phương pháp, thực nghiệm,	Nội dung: thiếu tối đa một thành phần bắt buộc.	Nội dung: thiếu tối đa hai thành phần bắt buộc.	Không làm

	<p>nghiêm, đánh giá, đánh giá, phân tích phân tích kết quả, kết quả, kết luận, và đưa ra nhận xét của bản thân.</p>	<p>đánh giá, phân tích kết quả, kết luận, và đưa ra nhận xét của bản thân.</p>			
	<p>Hình thức trình bày slide: Không vi phạm các qui tắc soạn slide đã được hướng dẫn tại lớp.</p>	<p>Hình thức trình bày slide : Vi phạm tối đa 01 qui tắc soạn slide đã được hướng dẫn tại lớp.</p>	<p>Hình thức trình bày slide: Vi phạm tối đa 02 qui tắc soạn slide đã được hướng dẫn tại lớp.</p>	<p>Hình thức trình bày slide: Vi phạm hơn 02 qui tắc soạn slide đã được hướng dẫn tại lớp.</p>	
	<p>Kỹ năng nói: trôi chảy, tự tin, thu hút và đảm bảo thời gian cho phép.</p>	<p>Kỹ năng nói: chưa thu hút nhưng trôi chảy, tự tin và thời gian báo cáo (<math>\pm 3</math> phút).</p>	<p>Kỹ năng nói: Chưa thu hút, tự tin nhưng trôi chảy và thời gian báo cáo (<math>\pm 3</math>-10 phút)</p>	<p>Kỹ năng nói: không chuẩn bị nghiêm túc.</p>	
	<p>Kỹ năng lắng nghe và trả lời câu hỏi: tham dự</p>	<p>Kỹ năng lắng nghe và trả lời câu hỏi: tham dự</p>	<p>Kỹ năng lắng nghe và trả lời câu hỏi: không tham</p>	<p>X</p>	

	thuyết trình, lắng nghe người đặt câu hỏi, trả lời hợp lý với thái độ phù hợp.	thuyết trình, lắng nghe người đặt câu hỏi với thái độ phù hợp.	dự, không lắng nghe hoặc không ứng xử tốt.		
--	--	--	--	--	--

b. Rubric của thành phần đánh giá A4

Bảng 8 - Tiêu chí đánh giá của thành phần A4 trên hình thức thi viết.

Tiêu chí	10	8	6	4	0
Hình thức	Trình bày sạch đẹp, rõ ràng, không sai chính tả	Trình bày rõ ràng, không sai chính tả	Trình bày sai chính tả, không rõ ràng, ...	Trình bày sai chính tả, không rõ ràng, ...	Không làm hoặc gian lận (sao chép).
Nội dung	Hoàn thành 100% nội dung	Hoàn thành 80% nội dung	Hoàn thành 60% nội dung	Hoàn thành 40% nội dung	

## 7. QUY ĐỊNH CỦA MÔN HỌC (Course requirements and expectations)

- Dự lớp: theo qui định chung của trường.
- Bài tập: Sinh viên phải làm các bài tập, phải chuẩn bị trả lời các câu hỏi và đọc trước những tài liệu theo yêu cầu của giảng viên trước mỗi buổi học.

## 8. TÀI LIỆU HỌC TẬP, THAM KHẢO

### Giáo trình

1. Kiet Van Nguyen. Khai phá dữ liệu truyền thông xã hội (Slides). Đại học Công nghệ Thông tin.
2. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu (2014). *Social Media Mining*. Cambridge University Press.

### Tài liệu tham khảo

1. Lei Tang, Huan Liu (2010). *Community Detection and Mining in Social Media*. Morgan & Claypool Publishers.

## 9. PHẦN MỀM HAY CÔNG CỤ HỖ TRỢ ĐỒ ÁN

7. Ngôn ngữ lập trình: Python.
8. Các thư viện máy học được hỗ trợ trên ngôn ngữ Python: nltk, sklearn, pytorch, ...

*Tp.HCM, ngày 02 tháng 5 năm 2022*

**Trưởng Khoa**

**Giảng viên biên soạn**

**Nguyễn Gia Tuấn Anh**

**Nguyễn Văn Kiệt**