

CHƯƠNG TRÌNH ĐÀO TẠO TRÌNH ĐỘ ĐẠI HỌC
NGÀNH ĐÀO TẠO: KHOA HỌC DỮ LIỆU VÀ TRÍ TUỆ NHÂN TẠO

TH94005: HỌC MÁY VÀ KHAI PHÁ DỮ LIỆU (MACHINE LEARNING AND DATA MINING)

I. Thông tin về học phần

- Học kỳ: 6
- Tín chỉ: **3 (Lý thuyết 3 - Thực hành 0 - Tự học: 9)**
- Giờ tín chỉ đối với các hoạt động học tập
 - + Học lý thuyết trên lớp: 30 tiết
 - + Làm bài tập trên lớp: 10 tiết
 - + Thuyết trình và thảo luận trên lớp: 5 tiết
- Tự học: 135 tiết (theo kế hoạch cá nhân hoặc hướng dẫn của giảng viên)
- Đơn vị phụ trách:
 - Bộ môn: Khoa học máy tính
 - Khoa: Công nghệ thông tin
- Học phần thuộc khối kiến thức:

Đại cương ☐		Cơ sở ngành ☐		Chuyên ngành ☒		Chuyên sâu ☐	
Bắt buộc ☐	Tự chọn ☐	Bắt buộc ☐	Tự chọn ☐	Bắt buộc ☒	Tự chọn ☐	Bắt buộc ☐	Tự chọn ☐

- Ngôn ngữ giảng dạy: tiếng Anh ☐ Tiếng Việt ☒

II. Chuẩn đầu ra của chương trình đào tạo và mục tiêu, kết quả học tập mong đợi của học phần

* Các chuẩn đầu ra và chỉ báo của chương trình đào tạo mà học phần đóng góp:

Chuẩn đầu ra Sau khi hoàn tất chương trình, sinh viên có thể:	Chỉ báo đánh giá việc thực hiện được chuẩn đầu ra
Kiến thức chuyên môn	
CĐR 1	1.3. Áp dụng sự hiểu biết về các vấn đề đương đại trong lĩnh vực Công nghệ thông tin.
Kiến thức chuyên môn	
CĐR 3	3.1 Xây dựng được hệ thống xử lý và phân tích dữ liệu.
CĐR 3	3.2. Xây dựng được hệ thống trí tuệ nhân tạo.
Kỹ năng chung	
CĐR 5.	5.1. ận dụng kỹ năng tư duy logic, kỹ năng phân biện để giải quyết các vấn đề chuyên môn trong lĩnh vực Công nghệ thông tin.
Kỹ năng chuyên môn	

Chuẩn đầu ra Sau khi hoàn tất chương trình, sinh viên có thể:	Chỉ báo đánh giá việc thực hiện được chuẩn đầu ra
CĐR 7	7.1. Phát triển hệ thống khai phá dữ liệu để giải quyết các vấn đề trong nhiều lĩnh vực.
Tự chủ và trách nhiệm	
CĐR 8	8.1. Thể hiện thói quen tự học, tự nghiên cứu, nâng cao trình độ chuyên môn.

*** Mục tiêu:**

- Về kiến thức: Học phần cung cấp cho sinh viên tổng quan kiến thức, các khái niệm về học máy và mạng học sâu. Đồng thời đi sâu tìm hiểu các mô hình mạng học sâu trong việc xử lý dữ liệu lớn.
- Về kỹ năng: Học phần rèn luyện cho sinh viên kỹ năng sử dụng một số công cụ trong việc phân tích và xử lý với dữ liệu lớn nhiều chiều.
- Năng lực tự chủ và trách nhiệm: Học phần rèn luyện cho sinh viên ý thức về trách nhiệm cá nhân đối với việc học tập, tuân thủ các yêu cầu, các quy định chung; Thái độ hợp tác trong nhóm; làm việc theo nhóm, nghiêm túc, chủ động tìm hiểu tích lũy kiến thức và nâng cao trình độ bản thân.

*** Kết quả học tập mong đợi của học phần:**

Học phần đóng góp cho Chuẩn đầu ra sau đây của CTĐT theo mức độ sau:

I – Giới thiệu (Introduction); P – Thực hiện (Practice); R – củng cố (Reinforce); M – Đạt được (Master)

Mã HP	Tên HP	Mức độ đóng góp của học phần cho chỉ báo CĐR của CTĐT					
		1.3	3.1	3.2	5.1	7.1	8.1
TH94005	Học máy và khai phá dữ liệu	P	R	M	R	R	P

Ký hiệu	KQHTMD của học phần Hoàn thành học phần này, sinh viên thực hiện được	CĐR của CTĐT
Kiến thức chung		
K1	Thiết lập được mô hình mạng học sâu theo bài toán yêu cầu	1.3. Áp dụng sự hiểu biết về các vấn đề đương đại trong lĩnh vực Công nghệ thông tin.
Kiến thức chuyên môn		
K2	Trình bày được kết quả của mô hình mạng học sâu với dữ liệu bài toán	3.1. Phân tích được các vấn đề cơ bản về công nghệ liên quan đến xây dựng hệ thống thông tin.
K3	Phân tích kết quả của phân tích dữ liệu lớn và đưa ra kết luận tối ưu cho từng trường hợp cụ thể.	3.2. Thiết kế hệ thống xử lý và phân tích dữ liệu (áp dụng với hướng chuyên sâu Khoa học máy tính).

Kỹ năng chung		
K4	Sử dụng thành thạo ngôn ngữ Python để phân tích xử lý một số bài toán về phân tích dữ liệu lớn.	5.1. Vận dụng kỹ năng tư duy logic, kỹ năng phân biện để giải quyết hoặc đánh giá các vấn đề chuyên môn.
Kỹ năng chuyên môn		
K5	Giải quyết các yêu cầu bài toán phân tích dữ liệu thực tế bằng các công cụ học máy và khai phá dữ liệu	7.1. Phát triển hệ thống xử lý và phân tích dữ liệu (áp dụng với hướng chuyên sâu Khoa học máy tính).
Tự chủ và trách nhiệm		
K6	Hình thành khả năng làm việc độc lập, cần cù, sáng tạo, chính xác, tinh thần trách nhiệm, hợp tác cao trong lập trình giải quyết các bài toán xử lý dữ liệu.	8.1. Thể hiện thói quen tự học, tự nghiên cứu, nâng cao trình độ chuyên môn.

III. Nội dung tóm tắt của học phần

TH94005. Học máy và khai phá dữ liệu (3TC: 3 –0 – 9). Mô tả vắn tắt nội dung: Học phần gồm 4 chương với nội dung: Các nguyên lý của khai phá dữ liệu; Học có giám sát; Học không giám sát và ứng dụng vào các bài toán dữ liệu lớn.

IV. Phương pháp giảng dạy và học tập

1. Phương pháp giảng dạy

- Thuyết giảng và giảng dạy thông qua thực hành trực tiếp trên giảng đường hoặc trực tuyến thông qua Ms Teams.

2. Phương pháp học tập

- Sinh viên nghe giảng, kết hợp với tự học và trao đổi với bạn học và thầy cô trên giảng đường hoặc trực tuyến thông qua Ms Teams.

V. Nhiệm vụ của sinh viên

- Chuyên cần: Theo quy định chung của Học viện.

- Chuẩn bị cho bài giảng: Tất cả sinh viên tham dự học phần này phải đọc sách tham khảo trước khi đến lớp.

- Bài tập: Tất cả sinh viên tham dự học phần này phải hoàn thành ít nhất 50% lượng bài tập được giao.

- Thi giữa kì: Tất cả các sinh viên tham dự học phần này đều phải tham gia thi giữa kỳ.

- Thi cuối kì: Tất cả các sinh viên tham dự học phần này đều phải tham gia thi cuối kỳ.

VI. Đánh giá và cho điểm

1. Thang điểm: 10

2. Điểm trung bình của học phần là tổng điểm của các rubric nhân với trọng số tương ứng của từng rubric:

- **Đánh giá quá trình:**

+ Chuyên cần: 10 %

+ Bài tập giữa kỳ: 30%

- **Đánh giá cuối kì:**

+ Thi cuối kì: 60%

Bảng 1: Kế hoạch đánh giá và trọng số

Hoạt động đánh giá	KQHTMĐ được đánh giá	Trọng số (%)	Thời gian/Tuần học
Đánh giá quá trình			
Tham dự lớp	K1-K6	10	1-15
Bài tập giữa kỳ	K1-K5	30	13
Đánh giá cuối kỳ			
Thi cuối kỳ	K1-K5	60	Theo lịch

Bảng 2. Chỉ báo thực hiện các kết quả học tập mong đợi của học phần
(Chỉ dùng cho đánh giá và thi giữa kỳ, cuối kỳ theo hình thức trắc nghiệm)

Nội dung kiểm tra	Chỉ báo thực hiện của học phần được đánh giá qua câu hỏi	KQHTMĐ của môn học được đánh giá qua câu hỏi
Kỹ thuật khai phá dữ liệu Bayesian, k-NN, Tree, Cart	Chỉ báo 1: Dự báo kết quả theo bài toán yêu cầu Chỉ báo 2: Trình bày được kết quả đánh giá dữ liệu theo bài toán yêu cầu	K1 K2
Ứng dụng các mô hình học có giám sát vào phân tích dữ liệu lớn.	Chỉ báo 3: Phân tích kết quả của phân tích dữ liệu và đưa ra kết luận tối ưu cho từng trường hợp cụ thể.	K3
Xử lý bài toán phân tích dữ liệu từ yêu cầu thực tế	Chỉ báo 4: Phân tích kết quả của phân tích dữ liệu lớn.	K4
Ứng dụng các mô hình học không có giám sát vào phân tích dữ liệu lớn	Chỉ báo 5: Cài đặt được mô hình phân tích dữ liệu phù hợp với bài toán yêu cầu	K5

4. Các yêu cầu, quy định đối với học phần

Nộp bài tập chậm: Tất cả các trường hợp nộp bài tập chậm đều không được chấp nhận.

Tham dự các bài thi: Không tham gia bài thi giữa kỳ sẽ không được thi cuối kỳ. Trong trường hợp có lý do chính đáng sẽ được giảng viên bố trí cho kiểm tra bù (**trừ bài thi kết thúc học phần**)

Yêu cầu về đạo đức: Sinh viên có thái độ học tập nghiêm túc.

VII. Giáo trình/ tài liệu tham khảo

* Sách giáo trình/Bài giảng:

- Andreas C. Müller & Sarah Guido, *Introduction to machine learning with python*, O'Reilly, 2017.
- Ian H. Witten, Eibe Frank and Mark A. Hal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann; 3rd edition.

* Tài liệu tham khảo khác:

- O'Reilly Radar Team(2012). *Big data now*, NXB O'Reilly.

VIII. Nội dung chi tiết của học phần

Tuần	Nội dung	KQHTMĐ của học phần
1-6	Chương 1: Các nguyên lý của khai phá dữ liệu	

Tuần	Nội dung	KQHTMD của học phần
	A/ Các nội dung chính trên lớp: (18 tiết) Nội dung GD lý thuyết: (18 tiết) 1.1 Giới thiệu khai phá dữ liệu 1.2 Khoảng cách, biến đổi dữ liệu 1.3 Trực quan hóa cấu trúc dữ liệu 1.4 Phân tích thành phần chính 1.5 Nguyên lý Bayesian 1.6 Thuật toán trong khai phá dữ liệu 1.7 Bài tập	K1
	B/ Các nội dung cần tự học ở nhà: (54 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1
7-12	Chương 2: Học có giám sát	
	A/ Tóm tắt các nội dung chính trên lớp: (18 tiết) Nội dung GD lý thuyết: (18 tiết) 2.1 k-láng giềng gần nhất (k-NN) 2.2 Cây quyết định 2.3 Bayesian 2.4 Hồi quy tuyến tính 2.5 Mạng nơ ron 2.6 Bài tập về học máy và xử lý dữ liệu	K2
	B/ Các nội dung cần tự học ở nhà: (54 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1, K2
13-14	Chương 3: Học không giám sát	
	A/ Tóm tắt các nội dung chính trên lớp: (6 tiết) Nội dung GD lý thuyết: (9 tiết) 3.1 Phương pháp KMeans 3.2 Mạng tích chập trích xuất đặc trưng	K2, K4, K5
	B/ Các nội dung cần tự học ở nhà: (27 tiết) Sinh viên làm các bài tập tương ứng các mục đã nghe giảng trên lớp	K1-K5
15	Ôn tập và kiểm tra giữa kỳ	

IX. Yêu cầu của giảng viên đối với học phần

- Phòng học: đủ chỗ ngồi, đủ ánh sáng.
- Phương tiện phục vụ giảng dạy: có máy chiếu, micro.
- Các phương tiện khác: bảng viết phấn, phấn, khăn lau bảng.
- Elearning: phần mềm dạy và học trực tuyến MS Teams, máy tính, hệ thống máy chủ và hạ tầng kết nối mạng Internet với băng thông đáp ứng nhu cầu người dùng, không để xảy ra nghẽn mạng hay quá tải. Phòng học trực tuyến đầy đủ ánh sáng, cách âm tốt, thông thoáng, ngăn nắp, gọn gàng, sạch sẽ.

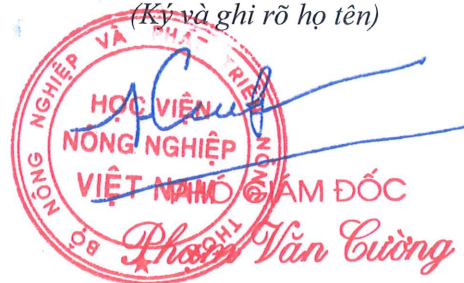
TRƯỞNG BỘ MÔN
(Ký và ghi rõ họ tên)


Trần Thị Thu Hằng
KT. TRƯỞNG KHOA
(Ký và ghi rõ họ tên)


Phạm Quang Dũng

Hà Nội, ngày...21...tháng...6...năm 2024
GIẢNG VIÊN BIÊN SOẠN
(Ký và ghi rõ họ tên)


Nguyễn Trung Cường
KT. GIÁM ĐỐC
(Ký và ghi rõ họ tên)



PHỤ LỤC

THÔNG TIN VỀ ĐỘI NGŨ GIẢNG VIÊN GIẢNG DẠY HỌC PHẦN

Giảng viên phụ trách học phần

Họ và tên: Nguyễn Trọng Kương	Học hàm, học vị: Tiến sĩ
Địa chỉ cơ quan: Khoa CNTT, Học viện Nông nghiệp Việt Nam	Điện thoại liên hệ:
Email: ntkuong@vnua.edu.vn	Trang web: fita.vnua.edu.vn/ntkuong
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên giảng dạy học phần

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên giảng dạy học phần

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

Giảng viên hỗ trợ /trợ giảng (nếu có)

Họ và tên:	Học hàm, học vị:
Địa chỉ cơ quan:	Điện thoại liên hệ:
Email:	Trang web: (Đưa tên website của Khoa; website cá nhân – nếu có)
Cách liên lạc với giảng viên: (nêu rõ hình thức liên lạc giữa sinh viên với giảng viên/trợ giảng)	

CÁC LẦN CẢI TIẾN (ĐỀ CƯƠNG ĐƯỢC CẢI TIẾN HÀNG NĂM THEO QUY ĐỊNH CỦA HỌC VIỆN)

- Lần 1: 07/2020

• Bổ sung phương pháp giảng dạy và học tập, nhiệm vụ của sinh viên, phương pháp đánh giá để phù hợp với việc triển khai dạy và học theo hình thức trực tuyến khi có dịch Covid-19.

- Lần 2: 07/2021

• Rà soát và cập nhật nội dung kiến thức học phần, rà soát tài liệu tham khảo

- Lần 3: 07/2022

- Rà soát và cập nhật tài liệu tham khảo

- Lần 4: 07/2023

- Rà soát và cập nhật mức đóng góp của học phần theo Chuẩn đầu ra và ma trận IPRM mới.
- Rà soát và cập nhật tài liệu tham khảo.
- Rà soát và cập nhật nội dung học phần, bài giảng theo tài liệu mới.

- Lần 5: 07/2024

- Cập nhật CDR của CTĐT
- Cập nhật đóng góp bảng IPRM
- Thay đổi mã môn học
- Rà soát và cập nhật nội dung học phần