

Introduction to Machine Learning and Data Mining

Phiên bản: 2021.3.05

1. THÔNG TIN CHUNG

Tên học phần:	Nhập môn học máy và khai phá dữ liệu (<i>Introduction to Machine Learning and Data Mining</i>)
Mã số học phần:	IT3190E
Code:	IT3190E
Khối lượng:	3(3-1-0-6)
Credit:	<ul style="list-style-type: none">- Lý thuyết - Lecture: 45 tiết- Bài tập - Exercise: 15 hours (If capstone project is used, please indicate clearly)- Thí nghiệm - Experiments: 0 hours
Học phần tiên quyết: Prerequisite	
Học phần học trước: Prior course	<ul style="list-style-type: none">- IT3312E (Data Structures and Algorithms)- IT3160E (Introduction to Artificial Intelligence)
Học phần song hành	Không
Paralell course:	No

2. MÔ TẢ HỌC PHẦN

Học phần này nhằm cung cấp cho sinh viên các kiến thức cơ bản về học máy và khai phá dữ liệu. Học phần sẽ trình bày quá trình học máy và khai phá dữ liệu cùng với các kỹ thuật cơ bản đối với các bài toán của học máy và khai phá dữ liệu. Các kỹ thuật này được cài đặt trong các thành phần lõi hệ thống để đưa ra các dự đoán và phát hiện tri thức trong các ứng dụng khác nhau. Học phần cũng giới thiệu các ứng dụng điển hình của học máy và khai phá dữ liệu trong thực tế và các công cụ cũng như thư viện hữu ích.

Ngoài ra học phần cũng cung cấp cho sinh viên các kỹ năng làm việc nhóm, thuyết trình và thái độ cần thiết để làm việc và ứng dụng các kỹ thuật học máy và khai phá dữ liệu trong các công ty sau này.

This course will provide the fundamental knowledge about Machine Learning (ML) and Data Mining (DM). It contains basic concepts, problems, and methods/models of ML/DM, as well as the methodology to work with data. Those methods/models can be implemented in different systems that can make prediction and discover new knowledge in real applications. The course also introduces some typical applications of ML and DM, and some useful libraries or frameworks. Beside, the course will enable the students to work in group, presentation skill, and the necessary attitude to work in the areas of ML and DM and in the future.

3. MỤC TIÊU VÀ CHUẨN ĐẦU RA CỦA HỌC PHẦN

Sinh viên hoàn thành học phần này có khả năng:

After this course the student will obtain the followings:

Mục tiêu/CĐ R Goal	Mô tả mục tiêu/Chuẩn đầu ra của học phần Description of the goal or output requirement	CDR được phân bổ cho HP/ Mức độ (I/T/U) Output division/ Level (I/T/U)
[1]	[2]	[3]
M1	Hiểu và có khả năng thiết kế và quản lý các hệ thống ứng dụng phương pháp của học máy và khai phá dữ liệu trong các tổ chức Understand and be able to design and manage machine learning and data mining methods in corporate information systems	1.1.1; 1.2.5; 1.1.4; 1.5.2; 2.3.3; 3.1.4
M1.1	Nhận diện và hiểu rõ các thành phần của hệ thống ứng dụng kỹ thuật học máy và khai phá dữ liệu Identify and understand the components of the systems that benefit from machine learning and data mining	[1.1.1; 1.2.5; 1.1.4] (I)
M1.2	Nhận diện, so sánh và phân loại được các nhu cầu trong doanh nghiệp mà có thể dùng các công nghệ của Học máy và khai phá dữ liệu để giải quyết Identify, compare, and categorize different data types and information systems in practice	[1.1.1; 1.2.5; 1.1.4] (T)
M1.3	Có khả năng thiết kế hệ thống ứng dụng kỹ thuật học máy và khai phá dữ liệu hỗ trợ xử lý dữ liệu, thông tin và tri thức trong tổ chức Be able to design systems which leverage machine learning and data mining to process data, information, and knowledge in their future organizations	[1.5.2; 2.3.3; 3.1.4] (TU)
M2	Nhận diện và làm chủ được các cơ hội trên thị trường do học máy và khai phá dữ liệu đem lại để phát triển tổ chức sẵn có và tạo ra các tổ chức mới Identify and manage the opportunities from machine learning and data mining to boost the existing organizations or develop new organizations	1.1.4; 1.3.3; 1.1.4; 1.3.3; 1.3.4; 1.6.3;
M2.1	Hiểu và vận dụng được các ứng dụng sử dụng học máy và khai phá dữ liệu đương đại nhằm hỗ trợ các hoạt động trong tổ chức Understand and use the recent advances from machine learning and data mining to support activities in their organizations	[1.1.4; 1.3.3; 1.6.3] (T)
M2.2	Nhận diện được các tác động của học máy và khai phá dữ liệu đối với tổ chức và môi trường hoạt động của tổ chức Identify the impacts of machine learning and data mining on their organizations	[1.3.4] (U)
M3	Nhận diện các xu hướng phát triển của học máy và khai phá dữ liệu có khả năng hỗ trợ việc thay đổi các	4.1.1; 4.1.2; 4.1.3

	tổ chức Identify the trends in machine learning and data mining that are able to support the development in organizations	
M3.1	Chủ động tìm hiểu và nhận diện các ứng dụng học máy và khai phá dữ liệu mới nhất Actively update and identify the most advancement in machine learning and data mining	[4.1.1; 4.1.2] (T)
M3.2	Xác định được các cơ hội tương lai mà học máy và khai phá dữ liệu đem lại để phát triển tổ chức sẵn có Identify the opportunities from machine learning and data mining to develop their organizations	[4.1.3] (U)

4. TÀI LIỆU HỌC TẬP

Reference book

- [1] Tom Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [2] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.
- [3] Jiawei Han, Micheline Kamber, Jian Pei. *Data Mining: Concepts and Techniques* (3rd Edition). Morgan Kaufmann, 2011.
- [4] Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. *Introduction to Data Mining* (2nd Edition). Pearson, 2017.
- [5] Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd Edition). Springer, 2009.

5. CÁCH ĐÁNH GIÁ HỌC PHẦN - EVALUATION

Điểm thành phần Module	Phương pháp đánh giá cụ thể Evaluation method	Mô tả Detail	CĐR được đánh giá Output	Tỷ trọng Percent
[1]	[2]	[3]	[4]	[5]
A1. Điểm quá trình Mid-term (*) (Một trong hai lựa chọn: A1.1 hoặc A1.2 Option A1.1 or A1.2)	Đánh giá quá trình Progress			50%
	A1.1. Bài tập về nhà Homework	Tự luận Written	M1.1; M1.2; M2.1; M2.2	20%
	A1.2. Bài tập lớn Capstone project	Báo cáo Presentation	M1.3; M3.1; M3.2	20%
A2. Điểm cuối kỳ Final term	A2.1. Thi cuối kỳ Final exam	Thi viết Written exam	M1÷M2	60%

* Điểm quá trình sẽ được điều chỉnh bằng cách cộng thêm điểm chuyên cần. Điểm chuyên cần có giá trị từ -2 đến +1, theo Quy chế Đào tạo đại học hệ chính quy của Trường ĐH Bách khoa Hà Nội.

The evaluation about the progress can be adjusted with some bonus. The bonus should belong to [-2, +1], according to the policy of Hanoi University of Science and Technology.

6. KẾ HOẠCH GIẢNG DẠY – SCHEDULE

Tuần Week	Nội dung Content	CDR học phần Output	Hoạt động dạy và học Activities	Bài đánh giá Evaluation
[1]	[2]	[3]	[4]	[5]
1	Chương 1: Giới thiệu chung về học máy và khai phá dữ liệu 1.1. Học máy và các bài toán học máy 1.2. Giới thiệu về khai phá dữ liệu 1.3. Các công cụ và thư viện của học máy và khai phá dữ liệu Chapter 1: Introduction to machine learning and data mining 1.1. Machine learning and some learning problems 1.2. Data mining 1.3. Some tools and libraries for machine learning and data mining	M1.1 M2.2 M3.2	Giảng bài Lecture	A1.2 A1.3
2	Chương 2: Thu thập và tiền xử lý dữ liệu 2.1. Thu thập dữ liệu 2.2. Trực quan hóa dữ liệu 2.3. Làm sạch dữ liệu 2.4. Tích hợp dữ liệu 2.5. Biến đổi dữ liệu 2.6. Lựa chọn đặc trưng Chapter 2: Data crawling and preprocessing 2.1. Data crawling 2.2. Data visualization 2.3. Data preprocessing 2.4. Data integration 2.5. Data transformation 2.6. Feature selection	M1.1 M1.2	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A1.2 A2.1
3	Chương 3: Hồi qui 3.1. Giới thiệu bài toán hồi qui 3.2. Mô hình tuyến tính Chapter 3: Linear regression 3.1. Introduction to regression 3.2. Linear model	M1.2; M1.3; M2.1; M2.2	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A2.1

4	Chương 4: Phân cụm 4.1. Giới thiệu bài toán phân cụm 4.2. Tiêu chuẩn đánh giá chất lượng 4.2. Phân cụm phân hoạch (K-mean và các biến thể) Chapter 4: Clustering 4.1. Introduction to clusering 4.2. Evaluation measures 4.3. Clustering by K-means and variants	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A1.2 A2.1
5	4.4. Phân cụm phân cấp 4.4. Hierarchical clustering	M1.2; M1.3; M2.1; M2.2	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A1.1 A1.2 A2.1
6	Chương 5: Phân loại 5.1. Giới thiệu bài toán phân loại 5.2. Lựa chọn tham số và đánh giá hiệu năng Chapter 5: Classification 5.1. Introduction to classification 5.2. Model selection and assessment	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A1.1 A1.2
7	5.3. Học dựa trên láng giềng gần nhất 5.3. Neighbor-based learning	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài Material reading; Lecture	A1.1 A2.1
8	5.4. Cây quyết định và rừng ngẫu nhiên 5.4. Decision tree and Random Forest	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài; Material reading; Lecture	A1.1 A2.1
9	5.5. Học dựa trên xác suất 5.5. Probabilistic models	M1.2; M1.3;	Giảng bài; Lecture	A2.1
10	5.6. Mạng nơ ron nhân tạo 5.6. Neural networks	M1.2; M1.3;	Giảng bài; Lecture	A1.2
11	5.7. Máy véc tơ hỗ trợ	M1.2; M1.3;	Giảng bài;	A1.3 A1.2

	5.7. Support vector machine	M2.1; M2.2	Lecture	
12	Chương 6: Khai phá tập mục thường xuyên và các luật kết hợp 6.1. Các tập mục thường xuyên và các luật kết hợp 6.2. Các độ đo đánh giá 6.3. Thuật toán Apriori 6.4. Các ứng dụng của luật kết hợp Chapter 6: Frequent itemsets and association rule mining 6.1. Frequent itemsets and association rules 6.2. Evaluation measures 6.3. Apriori algorithm 6.4. Some applications of association rules	M3.1; M3.2	Giảng bài; Lecture	A2.1 A1.2
13	Chương 7: Thảo luận ứng dụng học máy và khai phá dữ liệu trong thực tế 7.1. Một số nhu cầu thực tế 7.2. Đưa nhu cầu thực tế về bài toán học máy và khai phá dữ liệu 7.3. Một số vấn đề và kinh nghiệm Chapter 7: Discussion about applications of machine learning and data mining 7.1. Some practical needs 7.2. Transformation of a practical need into a machine learning and data mining problem 7.3. Some experiences	M3.1; M3.2	Giảng bài; Trao đổi, chia sẻ kinh nghiệm Lecture; Discussion	A1.2
14	Báo cáo bài tập nhóm Capstone project report	M3.1; M3.2	Báo cáo bài tập nhóm Capstone project report	A1.2
15	Tổng kết và ôn tập <i>Summary</i>			

7. QUY ĐỊNH CỦA HỌC PHẦN - COURSE REQUIREMENT

(Các quy định của học phần nếu có)

8. NGÀY PHÊ DUYỆT - DATE:

Chủ tịch hội đồng

Nhóm xây dựng đề cương

Committee chair**Course preparation group**

Thân Quang Khoát
Nguyễn Nhật Quang
Nguyễn Thị Kim Anh
Phạm Văn Hải
Ngô Văn Linh
Đinh Viết Sang
Trịnh Anh Phúc
Nguyễn Duy Hiệp
Nguyễn Hồng Quang
Nguyễn Thị Hoàng Lan
Nguyễn Linh Giang
Đặng Tuấn Linh
Visani Muriel

9. QUÁ TRÌNH CẬP NHẬT - UPDATE INFORMATION

ST T No	Nội dung điều chỉnh Content of the update	Ngày tháng được phê duyet Date accepted	Áp dụng từ kỳ/ khóa A pplicable from	Ghi chú Note
1			
2			