

Đại Học Quốc Gia TP.HCM
Trường Đại Học Bách Khoa
Khoa Khoa học và Kỹ thuật Máy tínhVietnam National University - HCMC
Ho Chi Minh City University of Technology
Faculty of Computer Science and Engineering

ĐỀ CƯƠNG HỌC PHẦN

Course Syllabus

1. Thông tin về học phần (Course information)

1.1. Thông tin tổng quan (General information)

- Tên học phần: **Khai phá dữ liệu**

Course title: **Data Mining**

- Mã học phần (Course ID): **CO3029**

- Số tín chỉ (Credits): **3 (ETCS: 6)**

- Học kỳ áp dụng (Applied from semester): **20221**

- Tổ chức học phần (Course format):

Hình thức học tập (Teaching/study type)	Số tiết/giờ (Hours)	Số tín chỉ (Credits)	Ghi chú (Notes)
Lý thuyết (LT) (Lectures)	30		
Thảo luận (ThL)/Thực hành tại lớp (TH) (Tutorial)	0		
Thí nghiệm (TNg)/Thực tập xưởng (TT) (Labs/Practices)	0		
Bài tập lớn (BTL)/Đồ án (ĐA) (Projects)	45		
Tự học (Self-study)	72.67		
Khác (Others)	0		
Tổng cộng (Total)	150	3	

- Tỷ lệ đánh giá và hình thức kiểm tra/thi (Evaluation form & ratio)

Hình thức đánh giá (Evaluation type)	Tỷ lệ (Ratio)	Hình thức (Format)	Thời gian (Duration)
Thảo luận (ThL)/Thực hành tại lớp (TH) (Tutorial)			
Thí nghiệm (Labs/Practices)			
Bài tập lớn (BTL)/Đồ án (ĐA) (Projects)	30%		
Kiểm tra (Midterm Exam)	20%	Báo cáo thí nghiệm (Experimental report)	50 phút (minutes)
Thi (Final Exam)	50%	Trắc nghiệm (Multiple choice (MCQ))	90 phút (minutes)
Tổng cộng (Total)	100%		



1.2. Điều kiện tiên quyết (*Prerequisites*)

HT/KN: Recommended, TQ: Prereq, SH: Coreq

Mã học phần (Course ID)	Tên học phần (Course title)	Tiên quyết (TQ)/song hành (SH) (Prerequisite - Prereq/Co - requisite - Coreq)
----------------------------	--------------------------------	--

1.3. Học phần thuộc khối kiến thức (*Knowledge block*)

- Kiến thức giáo dục đại cương (*General education*)
- Kiến thức giáo dục chuyên nghiệp (*Professional education*)
 - Kiến thức cơ sở ngành (*Foundation*) ☒ ◦ Kiến thức ngành (*Major*)
 - Kiến thức chuyên ngành (*Specialty*) ☒ ◦ Kiến thức Tốt nghiệp (*Graduation*)

1.4. Đơn vị phụ trách (Khoa/Bộ môn) (*Unit in-charge*)

Bộ môn / Khoa phụ trách (<i>Department</i>)	Hệ Thống Thông Tin - Khoa Khoa học và Kỹ thuật Máy tính (<i>Faculty of Computer Science and Engineering</i>)
Văn phòng (<i>Office</i>)	Khoa Khoa học và Kỹ thuật Máy tính
Điện thoại (<i>Phone number</i>)	5847
Giảng viên phụ trách (<i>Lecturer in-charge</i>)	Trần Minh Quang
E-mail	quangtran@hcmut.edu.vn

2. Mô tả học phần (*Course description*)

Môn học này nhằm giới thiệu quá trình khám phá tri thức, các khái niệm, công nghệ, và ứng dụng của khai phá dữ liệu. Ngoài ra, môn học này cũng trình bày các vấn đề tiền xử lý dữ liệu, các tác vụ khai phá dữ liệu, các giải thuật và công cụ khai phá dữ liệu mà có thể được dùng hỗ trợ nhà phân tích dữ liệu và nhà phát triển ứng dụng khai phá dữ liệu. Các chủ đề cụ thể của môn học bao gồm: tổng quan về khai phá dữ liệu, các vấn đề về dữ liệu được khai phá, các vấn đề tiền xử lý dữ liệu, hồi qui dữ liệu, phân loại dữ liệu, gom cụm dữ liệu, khai phá luật kết hợp, phát triển ứng dụng khai phá dữ liệu, và các đề tài nghiên cứu nâng cao trong khai phá dữ liệu.

This course aims to introduce the knowledge discovery process as well as concepts, technologies, and applications of data mining. It is also to discuss data preprocessing issues, data mining tasks, algorithms and tools that can be used to support data analysts and data mining application developers. In particular, its major topics are an overall view about data mining, issues related to data which are going to be mined, data preprocessing issues, data regression, data classification, data clustering, association rules mining, data mining application development, and other research topics of interest in the data mining area.

3. Giáo trình và tài liệu học tập (*Course materials*)

Sách, Giáo trình chính:

- [1] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining: Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers, 2012.
- [2] Trần Minh Quang, "Khai Phá Dữ Liệu và Kỹ Thuật Phân Lớp", NXB Đại Học Quốc Gia TP. HCM, 2020.

Sách tham khảo:

- [3] David Hand, Heikki Mannila, Padhraic Smyth, "Principles of Data Mining", MIT Press, 2001.
- [4] David L. Olson, Dursun Delen, "Advanced Data Mining Techniques", Springer-Verlag, 2008.
- [5] Graham J. Williams, Simeon J. Simoff, "Data Mining: Theory, Methodology, Techniques, and Applications", Springer-Verlag, 2006.
- [6] ZhaoHui Tang, Jamie MacLennan, "Data Mining with SQL Server 2005", Wiley Publishing, 2005.
- [7] Oracle, "Data Mining Concepts", B28129-01, 2008.
- [8] Oracle, "Data Mining Application Developer's Guide", B28131-01, 2008.
- [9] Ian H. Witten, Frank Eibe, Mark A. Hall, "Data mining: practical machine learning tools and techniques", Third Edition, Elsevier Inc, 2011.
- [10] Florent Messegli, Pascal Poncelet & Maguelonne Teisseire, "Successes and new directions in data mining", IGI Global, 2008.
- [11] Oded Maimon, Lior Rokach, "Data Mining and Knowledge Discovery Handbook", Second Edition, Springer Science + Business Media, LLC 2005, 2010.

Text books

- [1] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining: Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers, 2012.



[2] Trần Minh Quang, *Khai Phá Dữ Liệu và Kỹ Thuật Phân Lớp*, NXB Đại Học Quốc Gia TP. HCM, 2020.

Reference books

- [3] David Hand, Heikki Mannila, Padhraic Smyth, "Principles of Data Mining", MIT Press, 2001.
- [4] David L. Olson, Dursun Delen, "Advanced Data Mining Techniques", Springer-Verlag, 2008.
- [5] Graham J. Williams, Simeon J. Simoff, "Data Mining: Theory, Methodology, Techniques, and Applications", Springer-Verlag, 2006.
- [6] ZhaoHui Tang, Jamie MacLennan, "Data Mining with SQL Server 2005", Wiley Publishing, 2005.
- [7] Oracle, "Data Mining Concepts", B28129-01, 2008.
- [8] Oracle, "Data Mining Application Developer's Guide", B28131-01, 2008.
- [9] Ian H. Witten, Frank Eibe, Mark A. Hall, "Data mining: practical machine learning tools and techniques", Third Edition, Elsevier Inc, 2011.
- [10] Florent Messegli, Pascal Poncelet & Maguelonne Teisseire, "Successes and new directions in data mining", IGI Global, 2008.
- [11] Oded Maimon, Lior Rokach, "Data Mining and Knowledge Discovery Handbook", Second Edition, Springer Science + Business Media, LLC 2005, 2010.

4. Mục tiêu và kết quả học tập mong đợi (Goals and Learning outcomes)

4.1. Mục tiêu của học phần (Course goals)

Sau khi học đạt môn học này, sinh viên có thể:

- Minh họa được các bước trong quá trình khám phá tri thức
- Mô tả các khái niệm cơ bản, công nghệ và ứng dụng của khai phá dữ liệu
- Giải thích các tác vụ khai phá dữ liệu phổ biến như hồi qui, phân loại, gom cụm, và khai phá luật kết hợp
- Nhận dạng được các vấn đề về dữ liệu trong giai đoạn tiền xử lý cho các tác vụ khai phá dữ liệu
- Sử dụng các giải thuật và công cụ khai phá dữ liệu để phát triển ứng dụng khai phá dữ liệu

Upon successful completion, students will be able to:

- Demonstrate the steps in the overall knowledge discovery process
- Describe basic concepts, technologies, and applications of data mining
- Explain popular data mining tasks including regression, classification, clustering, and frequent itemset and association rules mining
- Identify data related issues in the data preprocessing phase for data mining tasks
- Use data mining algorithms and tools for data mining application development

4.2. Chuẩn đầu ra học phần (Course learning outcomes)

L.O.1 - Minh họa được các bước trong quá trình khám phá tri thức

(Demonstrate the steps in the overall knowledge discovery process)

L.O.1.1 - So sánh quá trình khám phá tri thức và quá trình khai phá dữ liệu

(Compare a knowledge discovery process with a data mining process)

L.O.1.2 - Liệt kê được các bước trong quá trình khám phá tri thức

(List the steps of a knowledge discovery process)

L.O.1.3 - Nêu ví dụ thực tế về quá trình khám phá tri thức

(Give a practical example of a knowledge discovery process)

L.O.2 - Mô tả các khái niệm cơ bản, công nghệ và ứng dụng của khai phá dữ liệu

(Describe basic concepts, technologies, and applications of data mining)

L.O.2.1 - Liệt kê các tác vụ khai phá dữ liệu

(List data mining tasks)

L.O.2.2 - Mô tả được các thành phần của tác vụ khai phá dữ liệu tổng quát

(Describe each component of a data mining task in general)

L.O.2.3 - Mô tả được các thành phần của giải thuật khai phá dữ liệu tổng quát

(Describe each component of a data mining algorithm in general)

L.O.2.4 - Mô tả được quy trình khai phá dữ liệu chuẩn

(Describe a standardized data mining process)

L.O.2.5 - Liệt kê được các ứng dụng của khai phá dữ liệu trong ít nhất 1 lĩnh vực thực tế

(List data mining applications in at least one application domain)



L.O.2.6 - Phân biệt được hệ thống khai phá dữ liệu với các dạng hệ thống khác như hệ cơ sở dữ liệu diễn dịch, hệ thống truy hỏi thông tin, hệ thống học máy, ...

(Determine the differences between a data mining system and other systems such as deductive database systems, information retrieval systems, machine learning systems, and so on)

L.O.3 - Giải thích các tác vụ khai phá dữ liệu phổ biến như hồi qui, phân loại, gom cụm, và khai phá tập mẫu thường xuyên và luật kết hợp

(Explain popular data mining tasks including regression, classification, clustering, and frequent itemset and association rules mining)

L.O.3.1 - Giải thích tác vụ hồi qui dữ liệu

(Explain data regression)

L.O.3.2 - Giải thích tác vụ phân loại dữ liệu

(Explain data classification)

L.O.3.3 - Giải thích tác vụ gom cụm dữ liệu

(Explain data clustering)

L.O.3.4 - Giải thích tác vụ khai phá tập mẫu thường xuyên và luật kết hợp

(Explain frequent itemset and association rules mining)

L.O.4 - Nhận dạng được các vấn đề về dữ liệu trong giai đoạn tiền xử lý cho các tác vụ khai phá dữ liệu

(Identify data related issues in the data preprocessing phase for data mining tasks)

L.O.4.1 - Xác định được các mô tả thống kê của tập dữ liệu cho trước

(Determine statistical descriptives of a given data set)

L.O.4.2 - Mô tả được vấn đề và giải pháp nhận diện nhiễu và phần tử ngoại biên trong tập dữ liệu cho trước

(Describe noise and outlier detection problems and solutions of a given data set)

L.O.4.3 - Thực hiện được các biến đổi dữ liệu trên tập dữ liệu cho trước

(Conduct data transformation on a given data set)

L.O.4.4 - Thực hiện được các thu giảm dữ liệu trên tập dữ liệu cho trước

(Conduct data reduction on a given data set)

L.O.5 - Sử dụng các giải thuật và công cụ khai phá dữ liệu để phát triển ứng dụng khai phá dữ liệu

(Use data mining algorithms and tools for data mining application development)

L.O.5.1 - Khai phá được mô hình hồi qui dữ liệu/mô hình phân loại dữ liệu/mô hình gom cụm dữ liệu/tập mẫu thường xuyên và luật kết hợp tương ứng trong ứng dụng khai phá dữ liệu

(Build data regression models/data classification models/data clustering models/frequent itemsets and association rules in data mining applications)

L.O.5.2 - Sử dụng được thư viện khai phá dữ liệu trong ứng dụng khai phá dữ liệu

(Utilize data mining libraries for data mining application development)

L.O.5.3 - Minh họa được việc sử dụng kết quả khai phá dữ liệu trong một chương trình ứng dụng cụ thể

(Demonstrate using mining models/patterns in a particular data mining application)

5. Phương thức giảng dạy và học tập (Teaching and assessment methods)

5.1. Phương thức giảng dạy (Teaching methods)

STT (No.)	Phương thức giảng dạy (Teaching methods)
1	Phương pháp học tập tích hợp (Blended learning)

5.2. Phương pháp giảng dạy (Teaching activities)

Loại hoạt động (Assessment methods)	Tên loại hoạt động (Components activities)	Nội dung (Content)



Loại hoạt động (Assessment methods)	Tên loại hoạt động (Components activities)	Nội dung (Content)
TES-Kiểm tra giữa kỳ (Midterm exam)	A.O.1 - Kiểm tra giữa kỳ (Midterm exam)	Tính 20% số điểm. Có thể được thay thế bằng bài trình bày trên lớp. (Account for 20% of the score. Can be conducted by in class presentations.)
EXM-Thi cuối kỳ (Final exam)	A.O.2 - Kiểm tra cuối kỳ (Final exam)	Tính 50% số điểm. Bài kiểm tra gồm các câu hỏi trắc nghiệm và trả lời ngắn. (Account for 50% of the score. The test consists of multi-choice questions and short writing questions.)
GHW-Bài tập nhóm về nhà (Group homework)	A.O.3 - Bài tập lớn/Tiểu luận (Assignment)	Tính 30% số điểm. Sinh viên thực hiện tiểu luận (mini project) theo nhóm từ 2 - 3 sinh viên theo chủ đề được cho trước ở tuần 1 hoặc tuần 2 (sinh viên có thể đề xuất chủ đề). Các nhóm báo cáo kết quả đồ án vào hai tuần cuối của học kỳ. (Account for 30% of the score. Students will work in group (2 - 3 students) for mini project. A topic will be assigned (student can propose a suitable topic) in the 1st or 2nd week. Students will present the mini project at the two last weeks.)
IHW-Bài tập cá nhân về nhà (Individual homework)	A.O.4 - Bài tập (Test/Quiz)	Các bài tập củng cố kiến thức tại lớp. Sinh viên thực hiện tốt sẽ được cộng điểm. (In-class quiz. Students can get bonus for good answers/solutions)

5.3. Hình thức đánh giá (Assessment methods)

Chuẩn đầu ra chi tiết (Learning outcome)	Hoạt động đánh giá (Evaluation activities)
L.O.1.1-So sánh quá trình khám phá tri thức và quá trình khai phá dữ liệu (Compare a knowledge discovery process with a data mining process)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.1.2-Liệt kê được các bước trong quá trình khám phá tri thức (List the steps of a knowledge discovery process)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.1.3-Nêu ví dụ thực tế về quá trình khám phá tri thức (Give a practical example of a knowledge discovery process)	A.O.2-Kiểm tra cuối kỳ (Final exam) A.O.3-Bài tập lớn/Tiểu luận (Assignment)
L.O.2.1-Liệt kê các tác vụ khai phá dữ liệu (List data mining tasks)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam) A.O.4-Bài tập (Test/Quiz)
L.O.2.2-Mô tả được các thành phần của tác vụ khai phá dữ liệu tổng quát (Describe each component of a data mining task in general)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.2.3-Mô tả được các thành phần của giải thuật khai phá dữ liệu tổng quát (Describe each component of a data mining algorithm in general)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.2.4-Mô tả được quy trình khai phá dữ liệu chuẩn (Describe a standardized data mining process)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.2.5-Liệt kê được các ứng dụng của khai phá dữ liệu trong ít nhất 1 lĩnh vực thực tế (List data mining applications in at least one application domain)	A.O.2-Kiểm tra cuối kỳ (Final exam) A.O.4-Bài tập (Test/Quiz)
L.O.2.6-Phân biệt được hệ thống khai phá dữ liệu với các dạng hệ thống khác như hệ cơ sở dữ liệu diễn dịch, hệ thống truy hồi thông tin, hệ thống học máy, ... (Determine the differences between a data mining system and other systems such as deductive database systems, information retrieval systems, machine learning systems, and so on)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)
L.O.3.1-Giải thích tác vụ hồi qui dữ liệu (Explain data regression)	A.O.1-Kiểm tra giữa kỳ (Midterm exam) A.O.2-Kiểm tra cuối kỳ (Final exam)



Chuẩn đầu ra chi tiết (<i>Learning outcome</i>)	Hoạt động đánh giá (<i>Evaluation activities</i>)
L.O.3.2-Giải thích tác vụ phân loại dữ liệu (<i>Explain data classification</i>)	A.O.1-Kiểm tra giữa kì (<i>Midterm exam</i>) A.O.2-Kiểm tra cuối kì (<i>Final exam</i>)
L.O.3.3-Giải thích tác vụ gom cụm dữ liệu (<i>Explain data clustering</i>)	A.O.1-Kiểm tra giữa kì (<i>Midterm exam</i>) A.O.2-Kiểm tra cuối kì (<i>Final exam</i>)
L.O.3.4-Giải thích tác vụ khai phá tập mẫu thường xuyên và luật kết hợp (<i>Explain frequent itemset and association rules mining</i>)	A.O.1-Kiểm tra giữa kì (<i>Midterm exam</i>) A.O.2-Kiểm tra cuối kì (<i>Final exam</i>)
L.O.4.1-Xác định được các mô tả thống kê của tập dữ liệu cho trước (<i>Determine statistical descriptives of a given data set</i>)	A.O.1-Kiểm tra giữa kì (<i>Midterm exam</i>) A.O.2-Kiểm tra cuối kì (<i>Final exam</i>)
L.O.4.2-Mô tả được vấn đề và giải pháp nhận diện nhiễu và phần tử ngoại biên trong tập dữ liệu cho trước (<i>Describe noise and outlier detection problems and solutions of a given data set</i>)	A.O.1-Kiểm tra giữa kì (<i>Midterm exam</i>) A.O.2-Kiểm tra cuối kì (<i>Final exam</i>)
L.O.4.3-Thực hiện được các biến đổi dữ liệu trên tập dữ liệu cho trước (<i>Conduct data transformation on a given data set</i>)	A.O.2-Kiểm tra cuối kì (<i>Final exam</i>) A.O.3-Bài tập lớn/Tiểu luận (<i>Assignment</i>)
L.O.4.4-Thực hiện được các thu giảm dữ liệu trên tập dữ liệu cho trước (<i>Conduct data reduction on a given data set</i>)	A.O.2-Kiểm tra cuối kì (<i>Final exam</i>) A.O.3-Bài tập lớn/Tiểu luận (<i>Assignment</i>)
L.O.5.1-Khai phá được mô hình hồi qui dữ liệu/mô hình phân loại dữ liệu/mô hình gom cụm dữ liệu/tập mẫu thường xuyên và luật kết hợp tương ứng trong ứng dụng khai phá dữ liệu (<i>Build data regression models/data classification models/data clustering models/frequent itemsets and association rules in data mining applications</i>)	A.O.2-Kiểm tra cuối kì (<i>Final exam</i>) A.O.3-Bài tập lớn/Tiểu luận (<i>Assignment</i>) A.O.4-Bài tập (<i>Test/Quiz</i>)
L.O.5.2-Sử dụng được thư viện khai phá dữ liệu trong ứng dụng khai phá dữ liệu (<i>Utilize data mining libraries for data mining application development</i>)	A.O.2-Kiểm tra cuối kì (<i>Final exam</i>) A.O.3-Bài tập lớn/Tiểu luận (<i>Assignment</i>)
L.O.5.3-Minh họa được việc sử dụng kết quả khai phá dữ liệu trong một chương trình ứng dụng cụ thể (<i>Demonstrate using mining models/patterns in a particular data mining application</i>)	A.O.2-Kiểm tra cuối kì (<i>Final exam</i>) A.O.3-Bài tập lớn/Tiểu luận (<i>Assignment</i>)

5.4. Hướng dẫn cách học (*Study guidelines*)

Để đáp ứng mục tiêu của môn học, sinh viên cần thực hiện tốt các đòi hỏi sau đây:

- Có mặt tại lớp phải hơn 75% từ tuần 1 đến tuần 8 và 100% từ tuần 9 đến tuần 15.
- Sau mỗi chương, sinh viên làm các bài tập của chương.

To meet the course objectives, students need to fulfill the following requirements:

- Attendance must be more than 75% from week 1 to week 8 and 100% from week 9 to week 15.
- After each chapter, students do the exercises of the chapter.

6. Nội dung chi tiết của học phần (*Course content*)

L.O. Chuẩn đầu ra chi tiết (*Detailed learning outcomes*)

A. Hoạt động đánh giá (*Assessment activity*)

Lec. Hoạt động dạy Giảng viên (*Lecturer*)

Stu. Hoạt động học Sinh viên (*Student*)

Buổi (<i>Session</i>)	Nội dung (<i>Content</i>)	Hoạt động dạy và học (<i>Lecturing</i>)
----------------------------	-----------------------------	---



Buổi (Session)	Nội dung (Content)	Hoạt động dạy và học (Lecturing)
1	<p>Chương 1: Tổng quan về khai phá dữ liệu</p> <ol style="list-style-type: none"> 1. Quá trình khám phá tri thức 2. Các khái niệm 3. Ý nghĩa và vai trò của khai phá dữ liệu 4. Ứng dụng của khai phá dữ liệu 5. Tóm tắt <p>Yêu cầu tự học đ/v sinh viên (6giờ)</p> <p><i>(Chapter 1: Overview of Data Mining</i> <i>1. Knowledge discovery process</i> <i>2. Concepts</i> <i>3. Meaning and role of data mining</i> <i>4. Applications of Data Mining</i> <i>5. Summary</i> <i>Self-study requirements for students (6 hours))</i></p>	<ul style="list-style-type: none"> • L.O.1.1 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make questions individually) • L.O.2.1 [A.O.4 , A.O.1] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make question individually) • L.O.2.2 [A.O.2 , A.O.1] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make question individually) • L.O.1.2 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make question individually) • L.O.1.3 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make question individually) • L.O.2.4 [A.O.2 , A.O.1] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi trên lớp theo cá nhân (make questions individually)
2	<p>Chương 2: Các vấn đề tiền xử lý dữ liệu</p> <ol style="list-style-type: none"> 3.1. Tổng quan về giai đoạn tiền xử lý dữ liệu 3.2. Tóm tắt mô tả về dữ liệu 3.3. Làm sạch dữ liệu 3.4. Tích hợp dữ liệu 3.5. Biến đổi dữ liệu 3.6. Thu giảm dữ liệu 3.7. Rời rạc hóa dữ liệu 3.8. Tạo cây phân cấp ý niệm 3.9. Biểu diễn dữ liệu 3.10. Tóm tắt <p>Yêu cầu tự học đ/v sinh viên (6 giờ)</p> <p><i>(Chapter 2: Data preprocessing issues</i> <i>3.1. Overview of the data preprocessing stage</i> <i>3.2. Descriptive summary of data</i> <i>3.3. Data cleaning</i> <i>3.4. Data integration</i> <i>3.5. Data Transformation</i> <i>3.6. Data collection</i> <i>3.7. Discrete data</i> <i>3.8. Create a hierarchy of ideas</i> <i>3.9. Data representation</i> <i>3.10. Summary</i> <i>Self-study requirements for students (6 hours))</i></p>	<ul style="list-style-type: none"> • L.O.4.4 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (Individual/group questions/exercises) • L.O.4.3 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (Individual/group questions/exercises) • L.O.4.1 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (Individual/group questions/exercises) • L.O.4.2 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (teaching) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (Individual/group questions/exercises)



Buổi (Session)	Nội dung (Content)	Hoạt động dạy và học (Lecturing)
3	<p>Chương 3: Hồi qui dữ liệu</p> <p>3.1. Tổng quan về hồi qui</p> <p>3.2. Hồi qui tuyến tính</p> <p>3.3. Hồi qui phi tuyến</p> <p>3.4. Ứng dụng</p> <p>3.5. Các vấn đề với hồi qui</p> <p>3.6. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (6 giờ)</p> <p><i>(Chapter 3: Data Regression</i> <i>3.1. Overview of regression</i> <i>3.2. Linear regression</i> <i>3.3. Nonlinear regression</i> <i>3.4. Application</i> <i>3.5. Problems with regression</i> <i>3.6. Summary</i> <i>Self-study requirements for students (6 hours))</i></p>	<ul style="list-style-type: none"> L.O.5.1 [A.O.4 , A.O.3] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) L.O.3.1 [A.O.1 , A.O.2] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)
4	<p>Chương 4: Phân loại dữ liệu</p> <p>4.1. Tổng quan về phân loại dữ liệu</p> <p>4.2. Phân loại dữ liệu với cây quyết định</p> <p>4.3. Phân loại dữ liệu với mạng Bayesian</p> <p>4.4. Phân loại dữ liệu với mạng Neural</p> <p>4.5. Các phương pháp phân loại dữ liệu khác</p> <p>4.6. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (12 giờ)</p> <p><i>(Chapter 4: Data Classification</i> <i>4.1. Overview of data classification</i> <i>4.2. Data classification with decision trees</i> <i>4.3. Data classification with Bayesian network</i> <i>4.4. Data classification with neural networks</i> <i>4.5. Other data classification methods</i> <i>4.6. Summary</i> <i>Self-study requirements for students (12 hours))</i></p>	<ul style="list-style-type: none"> L.O.3.2 [A.O.2 , A.O.1] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) L.O.5.1 [A.O.4 , A.O.3] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)
7	<p>Chương 5: Góm cụm dữ liệu</p> <p>5.1. Tổng quan về góm cụm dữ liệu</p> <p>5.2. Góm cụm dữ liệu bằng phân hoạch</p> <p>5.3. Góm cụm dữ liệu bằng phân cấp</p> <p>5.4. Góm cụm dữ liệu dựa trên mật độ</p> <p>5.5. Góm cụm dữ liệu dựa trên mô hình</p> <p>5.6. Các phương pháp góm cụm dữ liệu khác</p> <p>5.7. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (12 giờ)</p> <p><i>(Chapter 5: Clustering data</i> <i>5.1. Clustering overview</i> <i>5.2. Cluster data by partitioning</i> <i>5.3. Hierarchical clustering of data</i> <i>5.4. Clustering data based on density</i> <i>5.5. Model-based clustering</i> <i>5.6. Other Clustering Methods</i> <i>5.7. Summary</i> <i>Self-study requirements for students (12 hours))</i></p>	<ul style="list-style-type: none"> L.O.3.3 [A.O.1 , A.O.2] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) L.O.5.1 [A.O.4 , A.O.3] <ul style="list-style-type: none"> Lec: - Giảng lý thuyết (<i>teaching</i>) Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)



Buổi (Session)	Nội dung (Content)	Hoạt động dạy và học (Lecturing)
9	<p>Chương 6: Khai phá luật kết hợp</p> <p>6.1. Tổng quan về khai phá luật kết hợp</p> <p>6.2. Biểu diễn luật kết hợp</p> <p>6.3. Khám phá các mẫu thường xuyên</p> <p>6.4. Khám phá các luật kết hợp từ các mẫu thường xuyên</p> <p>6.5. Khám phá các luật kết hợp dựa trên ràng buộc</p> <p>6.6. Phân tích tương quan</p> <p>6.7. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (12 giờ)</p> <p><i>(Chapter 6: Exploring association rules</i> <i>6.1. An overview of association rule mining</i> <i>6.2. Representation of association rule</i> <i>6.3. Discover regular patterns</i> <i>6.4. Discover association rules from regular patterns</i> <i>6.5. Exploring constraint-based association rules</i> <i>6.6. Correlation analysis</i> <i>6.7. Summary</i> <i>Self-study requirements for students (12 hours))</i></p>	<ul style="list-style-type: none"> • L.O.5.1 [A.O.3 , A.O.4] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) • L.O.3.4 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)
11	<p>Chương 8: Các đề tài nghiên cứu trong khai phá dữ liệu</p> <p>8.1. Hướng dữ liệu</p> <p>8.2. Hướng kỹ thuật</p> <p>8.3. Hướng ứng dụng</p> <p>8.4. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (6 giờ)</p> <p><i>(Chapter 8: Research topics in data mining</i> <i>8.1. Data Direction</i> <i>8.2. Technical direction</i> <i>8.3. Application direction</i> <i>8.4. Summary</i> <i>Self-study requirements for students (6 hours))</i></p>	<ul style="list-style-type: none"> • L.O.2.1 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) • L.O.2.5 [A.O.4 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) • L.O.2.3 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: Hướng dẫn Làm bài tập (<i>homework guide</i>) ◦ Stu: Làm bài tập (<i>Do homework</i>) • L.O.2.6 [A.O.1 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)
11	<p>Chương 7: Phát triển ứng dụng khai phá dữ liệu</p> <p>7.1. Tổng quan về vấn đề phát triển ứng dụng khai phá dữ liệu</p> <p>7.2. Quy trình phát triển ứng dụng khai phá dữ liệu</p> <p>7.3. Các chuẩn dành cho khai phá dữ liệu</p> <p>7.4. Các công cụ hỗ trợ phát triển ứng dụng khai phá dữ liệu</p> <p>7.5. Tóm tắt</p> <p>Yêu cầu tự học đ/v sinh viên (6 giờ)</p> <p><i>(Chapter 7: Developing a Data Mining Application</i> <i>7.1. Overview of the problem of data mining application development</i> <i>7.2. Data mining application development process</i> <i>7.3. Standards for data mining</i> <i>7.4. Tools to support data mining application development</i> <i>7.5. Summary</i> <i>Self-study requirements for students (6 hours))</i></p>	<ul style="list-style-type: none"> • L.O.5.2 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) • L.O.5.3 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>) • L.O.5.1 [A.O.3 , A.O.2] <ul style="list-style-type: none"> ◦ Lec: - Giảng lý thuyết (<i>teaching</i>) ◦ Stu: - Câu hỏi/bài tập trên lớp theo cá nhân/nhóm (<i>Individual/group questions/exercises</i>)



Buổi (Session)	Nội dung (Content)	Hoạt động dạy và học (Lecturing)
12	<p>Báo cáo nhóm của sinh viên về các đề tài bài tập lớn của môn học</p> <p>Yêu cầu tự học đ/v sinh viên (24 giờ)</p> <p><i>(Student group reports on major subject matter topics)</i></p> <p><i>Request for self-study for students (24 hours))</i></p>	<ul style="list-style-type: none"> L.O.5.2 [A.O.3 , A.O.2] <ul style="list-style-type: none"> Lec: Chưa có (none) Stu: - Bài tập lớn theo nhóm - Báo cáo nhóm trên lớp (- Assignment - Report in class) L.O.5.3 [A.O.3 , A.O.2] <ul style="list-style-type: none"> Lec: Chưa có (none) Stu: - Bài tập lớn theo nhóm - Báo cáo nhóm trên lớp (- Assignment - Report in class) L.O.5.1 [A.O.3 , A.O.2] <ul style="list-style-type: none"> Lec: Chưa có (none) Stu: - Bài tập lớn theo nhóm - Báo cáo nhóm trên lớp (- Assignment - Report in class)

7. Yêu cầu khác về học phần (Other course requirements and expectations)

8. Biên soạn và cập nhật đề cương (Editing information)

- Đề cương được biên soạn vào năm học học kỳ (Syllabus edited in year-semester): **20221**
- Đề cương được chỉnh sửa lần thứ (Editing version): **DCMH.CO3029.3.1**
- Nội dung được chỉnh sửa, cập nhật, thay đổi ở lần gần nhất (The latest editing content): -- --

TRƯỞNG KHOA
(Dean)

CHỦ NHIỆM BỘ MÔN
(Head of Department)

Tp.Hồ Chí Minh, ngày 4 tháng 9 năm 2022
HCM City, September 4 2022
CB PHỤ TRÁCH LẬP ĐỀ CƯƠNG
(Lecturer in-charge)