

**1. THÔNG TIN CHUNG  
GENERAL INFORMATION**

<b>Tên học phần</b>	Introduction to Data Science
<b>Course name:</b>	
<b>Mã học phần</b>	IT4142E
<b>Code:</b>	
<b>Khối lượng</b>	2(2-1-0-6)
<b>Credit:</b>	<ul style="list-style-type: none"><li>- Lý thuyết - Lecture: 30 hours</li><li>- Bài tập – Exercise/Capstone Project: 15 hours</li><li>- Thí nghiệm - Experiments: 0 hours</li></ul>
<b>Học phần tiên quyết</b>	No
<b>Prerequisite:</b>	
<b>Học phần học trước</b>	<ul style="list-style-type: none"><li>- IT1110E: Introduction to Programming</li><li>- IT3090E: Database</li><li>- IT3010E: Data structures and algorithms</li><li>- IT3190E: Machine Learning</li></ul>
<b>Prior course:</b>	
<b>Học phần song hành</b>	No
<b>Paralell course:</b>	

**2. MÔ TẢ HỌC PHẦN - COURSE DESCRIPTION**

Học phần này nhằm cung cấp cho sinh viên các kiến thức cơ bản về Khoa học dữ liệu, một lĩnh vực liên ngành về các phương pháp, các quá trình, và các hệ thống có khả năng học/phát hiện tri thức từ dữ liệu. Các phương pháp và mô hình trong Khoa học dữ liệu sẽ giúp con người/máy đưa ra các quyết định và phán đoán tốt trong thực tế. Môn học sẽ giúp sinh viên nắm được và vận dụng được các bước chính khi phân tích dữ liệu, bao gồm tạo giả thuyết, lấy dữ liệu, tiền xử lý, phân tích, đánh giá chất lượng, và đưa ra phán đoán. Các phương pháp/mô hình từ Học máy (Machine Learning), Khai phá dữ liệu (Data Mining), và Thống kê (Statistics) sẽ được giới thiệu. Sinh viên sẽ được hướng dẫn làm sao có thể làm việc với dữ liệu text, image, videos, graphs, feedbacks,... Ngoài ra, môn học sẽ giới thiệu các công cụ và thư viện mà được ưa dùng trong thực tiễn.

This course introduces students to the field of Data Science, an interdisciplinary field of scientific methods, processes, and systems to extract knowledge from data. Methods from Data Science would support decision making and prediction. This course presents the key steps of data science processes, such as making assumption, data crawling, preprocessing, data analysis, knowledge evaluation, making prediction. Necessary methods from machine learning, data mining, and statistics will be introduced. The students will be introduced to how to work with texts, images, videos, graphs, social networks, ratings, feedbacks, ... This course also introduces the typical applications in practice and useful tools and libraries.

**3. MỤC TIÊU VÀ CHUẨN ĐẦU RA CỦA HỌC PHẦN**

## GOAL AND OUTPUT REQUIREMENT

Sinh viên hoàn thành học phần này có khả năng

After this course the student will obtain the followings:

<b>Mục tiêu/CDR Goal</b>	<b>Mô tả mục tiêu/Chuẩn đầu ra của học phần Description of the goal or output requirement</b>	<b>CDR được phân bổ cho HP/ Mức độ (I/T/U) Output division/ Level (I/T/U)</b>
<b>[1]</b>	<b>[2]</b>	<b>[3]</b>
<b>M1</b>	<b>Hiểu và có khả năng thiết kế và quản lý các hệ thống ứng dụng kỹ thuật của Khoa học dữ liệu trong các tổ chức</b> <b>Understand and be able to design and manage the systems which are based on Data Science</b>	[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 của Khoa học dữ liệu Identify and understand the components of the systems which are based on Data Science	[1.1.1; 1.2.5; 1.1.4] (I)
M1.2	Nhận diện, so sánh và phân loại được các dạng dữ liệu và hệ thống trong doanh nghiệp Identify, compare, and categorize the data type and 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 của Khoa học dữ liệu hỗ trợ xử lý dữ liệu, thông tin và tri thức trong tổ chức Be able to design systems which are based on Data Science to process data, information, and knowledge in their future organizations	[1.5.2; 2.3.3; 3.1.4] (TU)
<b>M2</b>	<b>Nhận diện và làm chủ được các cơ hội trên thị trường do Khoa học dữ liệu đem lại để phát triển tổ chức sẵn có và tạo ra các tổ chức mới</b> <b>Identify and manage the opportunities from Data Science to boost the existing organizations or develop new organizations</b>	[1.1.4; 1.3.3; 1.1.4; 1.3.3; 1.3.4]
M2.1	Hiểu và vận dụng được các ứng dụng của Khoa học dữ liệu đương đại nhằm hỗ trợ các hoạt động trong tổ chức Understand and use the recent advances from Data Science to support activities in their organizations	[1.1.4; 1.3.3] (T)
M2.2	Nhận diện được các tác động của Khoa học dữ liệu đối với tổ chức và môi trường hoạt động của tổ chức Identify the impacts of Data Science on their organizations	[1.3.4] (U)
<b>M3</b>	<b>Nhận diện các xu hướng phát triển của Khoa học dữ liệu có khả năng hỗ trợ việc thay đổi các tổ chức</b> <b>Identify the trends in Data Science that are able to support development in organizations</b>	[4.1.1; 4.1.2; 4.1.3]

M3.1	Chủ động tìm hiểu và nhận diện các ứng dụng của Khoa học dữ liệu mới nhất Actively update and identify the most advances in Data Science	[4.1.1; 4.1.2] (T)
M3.2	Xác định được các cơ hội mà Khoa học dữ liệu đem lại để phát triển tổ chức sẵn có Identify the opportunities from Data Science to develop their organizations	[4.1.3] (U)

#### 4. TÀI LIỆU HỌC TẬP

##### Reference

##### Reference book

- [1] Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques (3rd Edition). Morgan Kaufmann, 2011.
- [2] Joel Grus. Data Science from Scratch: First Principles with Python. O'Reilly, 2015.
- [3] Alberto Boschetti, Luca Massaron. Python Data Science Essentials (2nd Edition). Packt Publishing, 2016.
- [4] Trevor Hastie,<sup>†</sup> Robert Tibshirani,<sup>†</sup> Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition), Springer Publisher.

#### 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]
<b>A1. Điểm quá trình Mid-term (*)</b>	<b>Đánh giá quá trình Progress</b>			<b>40%</b>
	Bài tập nhóm Capstone Project	Báo cáo Presentation	M1.3; M3.1; M3.2	40%
<b>A2. Điểm cuối kỳ Final term</b>	<b>Thi cuối kỳ Final exam</b>	Trắc nghiệm hoặc Thi viết  Multiple-choice or Written exam	M1÷M2	<b>60%</b>

\* 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	CĐR học phần Output	Hoạt động dạy và học Teaching activities	Bài đánh giá Evaluate d in
[1]	[2]	[3]	[4]	[5]
1	<b>Giới thiệu chung</b> <ul style="list-style-type: none"> <li>- Giới thiệu</li> <li>- Cơ hội và thách thức</li> </ul> <p>Overview</p> <ul style="list-style-type: none"> <li>- Introduction</li> <li>- Opportunities and Challenges</li> </ul>	M1.1	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
2	<b>Thu thập và tiền xử lý dữ liệu</b> <ul style="list-style-type: none"> <li>- Giới thiệu máy thu thập dữ liệu</li> <li>- Các vấn đề tiền xử lý dữ liệu</li> <li>- Công cụ tiền xử lý dữ liệu</li> </ul> <p>Data crawling and preprocessing</p> <ul style="list-style-type: none"> <li>- Introduction to data crawler</li> <li>- Data preprocessing</li> <li>- Preprocessing tools</li> </ul>	M1.1 M1.2	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
3	<b>Làm sạch và tích hợp dữ liệu</b> <ul style="list-style-type: none"> <li>- Vấn đề chất lượng dữ liệu</li> <li>- Quy trình làm sạch dữ liệu</li> <li>- Tích hợp dữ liệu</li> </ul> <p>Data cleaning and integration</p> <ul style="list-style-type: none"> <li>- Data quality</li> <li>- Cleaning</li> <li>- Integration</li> </ul>	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
4	<b>Phân tích và khám phá dữ liệu</b> <p>Exploratory Data Analysis</p>	M1.2; M1.3; M2.1; M2.2	Đọc trước tài liệu; Giảng bài; Note reading; Teaching	A1 A2
5	<b>Trực quan hoá dữ liệu</b> <ul style="list-style-type: none"> <li>- Mục đích và kiểu dữ liệu</li> <li>- Nguyên lý thiết kế</li> <li>- Trực quan hoá bằng đồ thị</li> </ul> <p>Data visualization</p>	M1.3; M2.1; M2.2	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2

	<ul style="list-style-type: none"> <li>- Goal and data types</li> <li>- Design principles</li> <li>- Visualization by graphs</li> </ul>			
6	<b>Trực quan hoá dữ liệu đa biến</b>  Multidimensional data visualization	M1.3; M2.1; M2.2	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
7	<b>Học máy</b> <ul style="list-style-type: none"> <li>- Giới thiệu về Học máy và ứng dụng trong phân tích dữ liệu</li> <li>- Các giải thuật cơ bản</li> </ul> Machine Learning <ul style="list-style-type: none"> <li>- Introduction to machine learning and application to data analysis</li> <li>- Some basic algorithms</li> </ul>	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
8	<b>Phân tích dữ liệu lớn</b> <ul style="list-style-type: none"> <li>- Hệ sinh thái Hadoop</li> <li>- Mô hình lập trình MapReduce</li> <li>- Phân tích dữ liệu lớn</li> </ul> Big data analysis <ul style="list-style-type: none"> <li>- Hadoop ecosystem</li> <li>- MapReduce programming</li> <li>- Big data analysis</li> </ul>	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
9	<b>Báo cáo tiến độ bài tập lớn và hướng dẫn</b>  Capstone Project guidance	M3.1; M3.2	Các sinh viên báo cáo tiến độ; Progress report, Discussion	A1
10-11	<b>Option 1: Phân tích dữ liệu văn bản</b> <ul style="list-style-type: none"> <li>- Tổng quan về XLNNTN</li> <li>- Tách từ</li> <li>- Gán nhãn từ loại</li> </ul> Text analysis <ul style="list-style-type: none"> <li>- Introduction to natural language processing</li> <li>- Tokenization</li> <li>- POS tagging</li> </ul>	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
	<b>Option 2: Phân tích dữ liệu ảnh và video</b> <ul style="list-style-type: none"> <li>- Các khái niệm cơ bản về ảnh /video</li> <li>- Một số kỹ thuật tiền xử lý ảnh</li> </ul> Image and video analysis	M1.2; M1.3;	Đọc trước tài liệu; Giảng bài;	A1 A2

	<ul style="list-style-type: none"> <li>- Basic concepts in digital image/video</li> <li>- Some methods for image preprocessing</li> </ul>		Note reading; Teaching	
	<b>Option 3: Phân tích đồ thị</b> <ul style="list-style-type: none"> <li>- Tổng quan về phân tích mạng xã hội</li> <li>- Xếp hạng đồ thị</li> <li>- Nhận diện cộng đồng</li> </ul> <p>Graph analysis</p> <ul style="list-style-type: none"> <li>- Overview of social network analysis</li> <li>- Graph ranking</li> <li>- Community detection</li> </ul>	M3.1; M3.2	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
12	<b>Đánh giá kết quả phân tích</b> <ul style="list-style-type: none"> <li>- Quy trình đánh giá</li> <li>- Thực hành với bài toán thực tế</li> </ul> <p>Evaluation of analysis results</p> <ul style="list-style-type: none"> <li>- Evaluation process</li> <li>- Exercise</li> </ul>	M3.1; M3.2	Đọc trước tài liệu; Giảng bài;  Note reading; Teaching	A1 A2
13	Báo cáo bài tập nhóm Capstone project presentation	M3.1; M3.2	Báo cáo bài tập nhóm Project presentation	A1
14	Báo cáo bài tập nhóm Capstone project presentation	M3.1; M3.2	Báo cáo bài tập nhóm Project presentation	A1
15	<b>Summary</b>		Trao đổi Discussion	A1

## 7. QUY ĐỊNH CỦA HỌC PHẦN - COURSE REQUIREMENT

(The specific requirements if any)

## 8. NGÀY PHÊ DUYỆT - DATE: .....

**Chủ tịch hội đồng**  
**Committee chair**

**Nhóm xây dựng đề cương**  
**Course preparation group**  
Than Quang Khoat  
Nguyen Thi Oanh  
Nguyen Kiem Hieu  
Bui Thi Mai Anh

Tran Viet Trung  
Pham Van Hai  
Vasani Muriel  
Ho Tu Bao

## 9. QUÁ TRÌNH CẬP NHẬT - UPDATE INFORMATION

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