

Trường Đại học Bách Khoa -
ĐHQG Tp.HCM
Khoa: Khoa Khoa học và Kỹ
thuật Máy tính
Khoa/Bộ môn quản lý MH: Khoa
học Máy tính

Tp.HCM, ngày tháng
năm

Đề cương môn học Sau đại học

**NỀN TẢNG LẬP TRÌNH CHO PHÂN TÍCH VÀ TRỰC QUAN DỮ
LIỆU
(PROGRAMMING FOUNDATION FOR DATA ANALYTICS AND
VISUALIZATION)**

Mã số MH: CO5243

Số tín chỉ:	Tc (LT.BT&TH.Tự Học): 3					TCHP:					
Số tiết	-Tổng:	75	LT:	30	BT:	0	TH:	0	ĐA:	BTL/TL:	45
Đánh giá:			Thi:		40%		Trắc nghiệm, điền chỗ trống, tính toán và lập trình				
			Bài tập:		20%		Lập trình và tính toán				
			Bài tập lớn:		40%		Thực hiện dự án				
- Môn tiên quyết:											
- Môn học trước:											
- Môn song hành:											
- CTĐT ngành (Mã ngành):	Khoa Học Máy Tính (8480101)										
- Ghi chú khác:											

1. Mục tiêu môn học:

Môn học này giúp người học có khả năng lập trình để hiểu, hiện thực, và sử dụng được các kỹ thuật trong khoa học dữ liệu. Cụ thể, môn học nhằm trang bị cho người học những khả năng sau:

* hiểu và sử dụng được các tính năng của ngôn ngữ lập trình Python;

* hiểu, sử dụng, đánh giá, và lựa chọn được các công nghệ để triển khai các bước trong phân tích dữ liệu;

- * hiểu và sử dụng được các kỹ thuật để hiển thị trực quan dữ liệu;
- * hiểu và lập trình được các bước trong phân tích dữ liệu và dữ liệu lớn.

Aims:

This course helps students to have programming skills to understand, implement, and use techniques in data science. Concretely, this course help students to own the following abilities:

- * use Python for developing programs;
- * understand, use, evaluate, and select appropriate technologies to carry out data analysis;
- * understand and use techniques for visualizing data;
- * understand and implement tasks for analyzing data and big data.

2. Nội dung tóm tắt môn học:

Môn học cung cấp đến người học những khối kiến thức và kỹ năng sau đây:

- * Kỹ thuật lập trình với ngôn ngữ Python;
- * Numpy và Tensor để biểu diễn dữ liệu trong quá trình phân tích;
- * Các kỹ thuật hiển thị trực quan dữ liệu;
- * Cách sử dụng các thư viện để làm đọc, xử lý, và phân tích các dạng dữ liệu khác nhau, với quy mô khác nhau;
- * Các bước trong phân tích và hiển thị trực quan dữ liệu.

Course outline:

This course provide to students the following knowledge and skills:

- * Programming fundamentals with Python;
- * Numpy and Tensor to store data during the analysis process;
- * Techniques for visualizing data;
- * Important libraries for loading, processing, and analyzing differerent kinds of data at different scales;
- * Important tasks for analyzing and visualizing data and big data.

3. Tài liệu học tập:

[Python]

[1] Mark Lutz, “Learning Python - Powerful Object-Oriented Programming”, 5th Edition, O'Reilly Media, 2013.

[2] Bill Lubanovic, “Introducing Python - Modern Computing in Simple Packages”, 1st Edition, O'Reilly Media, 2015.

[Data Science: Programming]

[3] Jake VanderPlas, Python Data Science Handbook - Essential tools for working with data”, 1st Editions (Revision), O'Reilly Media, 2017.

[4] Aurélien Géron, “Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”, 2nd Edition, O'Reilly Media, 2019.

[5] Wes McKinney, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython”, 2nd Edition, O'Reilly Media, 2017.

[Visualization]

[6] Cole Nussbaumer Knaflic, “Storytelling with Data: A Data Visualization Guide for Business Professionals”, 1st Edition, Wiley, 2015.

[7] Stephanie D. H. Evergreen, “Effective Data Visualization: The Right Chart for the Right Data”, 1st Edition, 2015.

[8] Kirthi Raman, “Mastering Python Data Visualization”, 1st Edition, Packt Publishing, 2015.

[Library]

[9] OpenCV Python Tutorials: https://docs.opencv.org/3.1.0/d6/d00/tutorial_py_root.html#gsc.tab=0

[10] Natural Language Toolkit: <https://www.nltk.org>

[11] LibROSA: <https://librosa.github.io/librosa/>

[12] pandas: powerful Python data analysis toolkit: <https://pandas.pydata.org/pandas-docs/stable/>

[13] Learning PySpark: <https://github.com/PacktPublishing/Learning-PySpark>

[14] CVXOPT: <https://cvxopt.org>

[15] Pytorch: [PyTorchhttps://pytorch.org/](https://pytorch.org/)

[16] Keras: <http://keras.io>

[17] scikit-learn: <http://scikit-learn.github.io/stable>

4. Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học:

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CĐR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.1	Hiểu và sử dụng được các tính năng của ngôn ngữ lập trình Python.	Thi, Bài tập, Bài tập lớn/Tiểu luận	c, d, e	c, d, e	1.3, 2.1, 2.3

STT	Chuẩn đầu ra môn học (CĐRMH)	Công cụ đánh giá CĐRMH	Đóng góp CDR Chương trình (CĐRCT)		
			Ứng dụng	Nghiên cứu	
CĐRMH.2	Hiểu, sử dụng, đánh giá, và lựa chọn được các công nghệ để triển khai các bước trong phân tích dữ liệu.	Bài tập, Bài tập lớn/Tiểu luận	c, d, e	c, d, e	1.3, 2.1, 2.3, 3.1
CĐRMH.3	Hiểu và lập trình được các kỹ thuật để hiển thị trực quan dữ liệu.	Bài tập, Bài tập lớn/Tiểu luận	c, d, e	c, d, e	1.3, 2.1, 2.3
CĐRMH.4	Hiểu và lập trình được các bước trong phân tích dữ liệu và dữ liệu lớn.	Bài tập, Bài tập lớn/Tiểu luận	c, d, e	c, d, e	1.3, 2.1, 2.4, 3.1

Learning outcomes:

No.	Course learning outcomes (CLO)	CLO assessment	Matching with PLO		
			Coursework	Research	
L.O.1	Use Python for developing programs	Final Exam, Exercises, Assignment	c, d, e	c, d, e	1.3, 2.1, 2.3
L.O.2	Understand, use, evaluate, and select appropriate technologies to carry out data analysis.	Exercises, Assignment	c, d, e	c, d, e	1.3, 2.1, 2.3, 3.1
L.O.3	Understand and use techniques for visualising data.	Exercises, Assignment	c, d, e	c, d, e	1.3, 2.1, 2.3
L.O.4	Understand and implement tasks for analyzing data and big data.	Exercises, Assignment	c, d, e	c, d, e	1.3, 2.1, 2.4, 3.1

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình ứng dụng:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1			✓	✓	✓						
CĐRMH.2			✓	✓	✓						
CĐRMH.3			✓	✓	✓						
CĐRMH.4			✓	✓	✓						

Bảng ánh xạ chuẩn đầu ra môn học và chuẩn đầu ra chương trình nghiên cứu:

Chuẩn đầu ra môn học (CĐRMH)	Chuẩn đầu ra của chương trình (CĐRCT)										
	a	b	c	d	e	f	g	h	i	j	k
CĐRMH.1			✓	✓	✓						
CĐRMH.2			✓	✓	✓						
CĐRMH.3			✓	✓	✓						
CĐRMH.4			✓	✓	✓						

5. Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Môn học bao gồm không chỉ kiến thức mà còn kỹ năng về các kỹ thuật và công nghệ được dùng để hiện thực các dự án phân tích dữ liệu. Do đó, **người học cần đọc tài liệu và rèn luyện lập trình thông qua các bài tập và bài tập lớn** (hình thức tiểu luận) mới đạt được chuẩn đầu ra của môn học này.

Tài liệu (các hướng dẫn, các trang hỏi-đáp, các video và ví dụ mẫu) về các nội dung trong môn học này rất nhiều trên mạng Internet. Do đó, ngoài những tài liệu tham khảo được đề cập ở đề cương này, người học nên tìm ra cái phù hợp với mình hơn trên mạng.

Hiện nay, để rèn luyện lập trình Python trong phân tích dữ liệu, các học viên có thể dùng Google Colab hay các công cụ trên nền Cloud tương tự để có thể lập trình được tại mọi nơi (chỉ cần kết nối Internet) và không cần trang bị nhiều về phần cứng, vì các nền tảng đó đã trang bị cả CPU và GPU đủ mạnh cho thử nghiệm.

Môn học này dự kiến có hình thức đánh giá như sau:

* Thi cuối kỳ (40%): bài thi gồm các hình thức: trắc nghiệm, điền vào chỗ trống, lập trình, đưa ra giải pháp cho vấn đề và biện luận.

* Bài tập (20%): làm bài tập lập trình theo yêu cầu (làm cá nhân), chia sẻ code cho giảng viên và cho lớp (qua Google Colab hay công cụ tương tự)

* Bài tập lớn/Tiểu luận (40%): thực hiện dự án lập trình theo nhóm (2-3 học viên/nhóm), nộp báo cáo, chia sẻ code và video minh họa. Báo cáo bài tập lớn/tiểu luận tập trung.

Learning strategies & Assessment Scheme:

This course is to provide not only programming knowledge but also knowledge and skills of programming, and important techniques and technologies used in data science. Therefore, learners are requested **to read materials and to practise programming through doing programming exercises and implementing assignments.**

This course has a plentiful amount of materials (including tutorials, questions and answers, video and examples) on the web. Learners are encouraged to search for suitable materials beside of using recommended references.

Learners are also encouraged to use Google Colab and similar cloud-based tools for practising programming tasks in this course. By using such of those tools, learners can do excerses and assignments anywhere without worry about the power of CPU and GPU.

The grading policy of this course is as follows (tentative):

* Final exam (40%): including quiz, filling blanks, progamming, decribing and discussing different solutions for given questions.

* Programming excersies (20%): implementing exercises, sharing codes to lecturers and other members in class (using Google Colab or similar tools)

* Assignment (40%): implementing projects in a group of 2 or 3 members; submitting reports, sharing code and result videos. Presenting the results in a common session.

6. Nội dung chi tiết:

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
1	Giới thiệu	1. Giới thiệu về môn học 2. Cơ bản về lập trình Python 2.1 Công cụ lập trình 2.2 Nhập và xuất dữ liệu cơ bản 2.3 Các kiểu dữ liệu, biến và biểu thức 2.4 Các cấu trúc điều khiển 2.5 Định nghĩa và sử dụng hàm 2.6 Comprehension và Generation	CĐRMH.1	[1], [2]
2-3	Lập trình hướng đối tượng với Python	1. Cơ bản với lớp và đối tượng 2. Tính khả kiến 3. Hàm khởi tạo 4. Hàm cho lớp và cho đối tượng 5. Các hàm đặc biệt 6. Thừa kế 7. Cơ chế xử ngoại lệ 8. Định nghĩa và sử dụng các Iterator 9. Tổ chức chương trình 10. Hệ thống các lớp trong các thư viện dành cho học máy 10.1 Tensorflow/Keras 10.2 Pytorch	CĐRMH.1	[1], [2]
4	Sử dụng Numpy	1. Numpy là gì và tại sao? 2. Tổ chức dữ liệu cho đối tượng Numpy 3. Tạo các đối tượng Numpy 3.1 Tạo các đối tượng Numpy cơ bản 3.2 Sinh số ngẫu nhiên 3.3 Tạo và sử dụng đối tượng Numpy với API từ scikit-learn 4. Phép toán reshape 5. Phép concatenate và split 6. Index và slice trong numpy 7. Broadcast với đối tượng numpy 8. Hoán đổi các chiều dữ liệu 9. Các phép tính toán cơ bản với Numpy 9.1 Phép toán số học 9.2 Phép luận lý 9.3 Phép toán ma trận 9.4 Phép toán trong thống kê 9.5 Phép toán khác	CĐRMH.1, CĐRMH.2, CĐRMH.4	[3],[4],[5]
5	Hiển thị trực quan dữ liệu	1. Giới thiệu về hiển thị trực quan dữ liệu 2. Giới thiệu về matplotlib, seaborn, và các thư viện trực quan dữ liệu 3. Hiển thị trực quan dữ liệu 1-chiều, 2-chiều, và địa lý 4. Cách tùy biến thông tin trên biểu đồ 5. Biểu đồ histogram và các phân phối dữ liệu 6. Biểu đồ các đường đồng mức 7. Các biểu đồ cho đánh giá và so sánh 7.1 Biểu đồ sai số 7.2 Biểu đồ đường biên phân lớp 7.3 Biểu đồ các độ đo đánh giá và so sánh khác 8. Hiển thị dữ liệu ma trận ảnh 9. Hiển thị trực quan dữ liệu nhiều chiều 9.1 Tọa độ song song 9.2 Thu giảm số chiều + PCA + t-SNE	CĐRMH.1, CĐRMH.3	[6],[7],[8]

Tuần/ Buổi	Chủ đề (chương)	Nội dung	Chuẩn đầu ra môn học	Tài liệu
6	Làm việc với các dạng dữ liệu cơ bản	1. Đọc, xử lý và hiển thị dữ liệu ảnh 2. Đọc, xử lý và hiển thị dữ liệu văn bản 3. Đọc, xử lý và hiển thị dữ liệu chuỗi thời gian và âm thanh	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	[9],[10], [11]
7	Làm việc với các thư viện quan trọng trong khoa học dữ liệu	1. Tạo, xử lý và hiển thị dữ liệu với scikit-learn 2. Tạo, xử lý và hiển thị dữ liệu với Pandas 3. Tạo, xử lý và hiển thị dữ liệu với PySpark	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	[17], [12], [13]
8	Làm việc với tensor nhiều chiều	1. Giới thiệu về tensor trong các thư viện học sâu 2. Chuyển đổi giữa Numpy và các tensor 3. Các phép toán cơ bản trên tensor 4. Sử dụng GPU để tính toán trên tensor	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	[15], [16]
9	Học từ dữ liệu	1. Tối ưu có ràng buộc với CVXOPT 2. Tối ưu không ràng buộc với Tensorflow/Keras và Pytorch 3. Một số kỹ thuật học từ dữ liệu có chọn lọc	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	[3],[4],[5], [14-16]
10	Tổng hợp các kỹ thuật cho bài toán cụ thể	1. Đọc, biến đổi và làm sạch dữ liệu 2. Chuẩn hóa dữ liệu 3. Tóm lược và mô tả dữ liệu 4. Làm giàu dữ liệu 5. Học từ dữ liệu 6. Rút trích thông tin và ý nghĩa trên dữ liệu 7. Hiển thị trực quan dữ liệu 8. Xây dựng các biểu đồ đánh giá và so sánh	CĐRMH.1, CĐRMH.2, CĐRMH.3, CĐRMH.4	[1-17]

7. Giảng viên tham gia giảng dạy:

CBGD
chính:

TS. Lê
Thành
Sách

CBGD
tham
gia:

TS.
Trần
Tuấn
Anh
TS.
Nguyễn
Hồ
Mẫn
Rạng
TS.
Nguyễn
Quang
Hùng

**XÁC NHẬN
CỦA HỘI
ĐỒNG XÂY
DỰNG
CHƯƠNG
TRÌNH ĐÀO
TẠO VÀ KHOA**

*Tp. Hồ Chí
Minh, ngày
..... tháng
..... năm*

.....
**GIẢNG
VIÊN
LẬP ĐỀ
CƯƠNG**

**TS. Phạm
Hoàng
Anh**