

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN

1. Tên học phần: Khoa học dữ liệu (Introducing Data Science)

Mã số học phần: 7080509

Số tín chỉ học phần: 2 (2-0-4), học trong 8 tuần

Số tiết học phần:

Lý thuyết: 20;	Bài tập: 10;	Thực hành: 0;
Thực tập: 0;	Đồ án: 0;	Tự học: 60;

2. Đơn vị quản lý học phần

2.1. Giảng viên giảng dạy:

Mã giảng viên

1. ThS Đặng Văn Nam	0805-08
2. ThS Nguyễn Thùy Dương	0805-03
3. TS Tạ Quang Chiểu	0802-04
4. TS Nguyễn Thị Mai Dung	0803-05
5. TS Nguyễn Hoàng Long	0803-11
6. ThS Ngô Thị Phương Thảo	0803-08

2.2. Bộ môn: Khoa học máy tính.

2.3. Khoa: Công nghệ thông tin.

3. Điều kiện học học phần

3.1. Môn học tiên quyết: Xác suất thống kê (7080120)

3.2. Môn học học trước: Nhập môn ngành công nghệ thông tin (7080514)

4. Mục tiêu của học phần

4.1. Kiến thức

4.1.1. Hiểu các khái niệm cơ bản về khoa học dữ liệu và vai trò của khoa học dữ liệu trong cuộc cách mạng công nghiệp 4.0.

4.1.2. Vận dụng được các bước trong quy trình khoa học dữ liệu;

4.1.3. Biết được các công cụ và ngôn ngữ lập trình trong khoa học dữ liệu

4.1.4. Áp dụng ngôn ngữ lập trình Python và các thư viện phổ biến trong khoa học dữ liệu.

4.2. Kỹ năng

4.2.1. Có khả năng lập trình giải quyết một số bài toán với ngôn ngữ Python

4.2.2. Sử dụng được các thư viện Python trong khoa học dữ liệu

5. Tóm tắt nội dung học phần

Học phần giới thiệu những khái niệm cơ bản về khoa học dữ liệu, xu thế phát triển về dữ liệu lớn hiện tại và tương lai. Giới thiệu tổng quan về các nền tảng và tiến trình khoa học dữ liệu. Ngôn ngữ lập trình Python trong khoa học dữ liệu. Khai thác và sử dụng một số thư viện Python trong khoa học dữ liệu như: Thư viện xử lý dữ liệu số (Numpy), Thư viện trực quan hóa dữ liệu (Matplotlib), Thư viện phân tích dữ liệu (Pandas); Thư viện các giải thuật học máy (Scikit Learn)

6. Cấu trúc nội dung học phần

Bảng 1. Nội dung học phần

6.1. Lý thuyết			
Đề mục	Nội dung	Số tiết	Mục tiêu
Chương 1	Mở đầu	4	
1.1	Vai trò của khoa học dữ liệu trong thực tế	4	4.1.1 4.1.2
1.2	Tổng quan về khoa học dữ liệu		
1.3	Dữ liệu lớn và khoa học dữ liệu		
1.4	Quy trình khoa học dữ liệu		
1.5	Các kiến thức, kỹ năng của một nhà khoa học dữ liệu		
Chương 2	Ngôn ngữ và công cụ lập trình trong khoa học dữ liệu	2	
2.1	Giới thiệu một số ngôn ngữ và công cụ lập trình trong khoa học dữ liệu	2	4.1.3 4.2.1
2.2	Ngôn ngữ lập trình Python trong khoa học dữ liệu		
2.3	Cài đặt và thiết lập môi trường lập trình		
2.4	Xây dựng ứng dụng đầu tiên với Python		
2.5	Bài tập – Chương 2	1	
Chương 3	Lập trình Python căn bản	16	
3.1	Biến, toán tử trong Python	8	4.1.4 4.2.1
3.2	Các kiểu dữ liệu cơ bản		
3.3	Cấu trúc điều khiển, vòng lặp		
3.4	Hàm, Lớp, Đối tượng trong Python		
3.5	Làm việc với tập tin		
3.6	Bài tập – Chương 3	8	
Chương 4	Một số thư viện Python trong khoa học dữ liệu	8	
4.1	Giới thiệu một số Package trong khoa học dữ liệu	6	4.1.4

4.2	Xử lý dữ liệu số với thư viện Numpy		4.2.2
4.3	Trực quan hóa dữ liệu với thư viện Matplotlib		
4.4	Phân tích dữ liệu với thư viện Pandas		
4.5	Học máy với thư viện Scikit Learn		
4.6	Bài tập – Chương 4	2	

7. Phương pháp giảng dạy

- Giảng dạy lý thuyết kết hợp với bài tập, bài thực hành;
- Sử dụng các ví dụ minh họa trực quan cho từng nội dung cụ thể
- Hướng dẫn sinh viên phương pháp tự đọc, tự nghiên cứu, tìm kiếm tài liệu.

8. Nhiệm vụ của sinh viên

Sinh viên phải thực hiện các nhiệm vụ sau:

- Có mặt tối thiểu 80% số tiết học lý thuyết.
- Hoàn thành đầy đủ 100% giờ thực hành/thí nghiệm/thực tập và có báo cáo kết quả.
- Hoàn thành đầy đủ các bài tập nhóm và được đánh giá kết quả thực hiện.
- Tham dự đầy đủ các kiểm tra giữa học kỳ.
- Tham dự thi kết thúc học phần (bắt buộc).
- Chủ động chuẩn bị các nội dung và thực hiện giờ tự học theo mục 11.

9. Đánh giá kết quả học tập của sinh viên

9.1. Cách đánh giá

Sinh viên được đánh giá tích lũy học phần như sau:

Bảng 2. Đánh giá học phần

TT	Điểm thành phần	Quy định	Trọng số
1	Điểm chuyên cần	Số tiết tham dự học/tổng số tiết	10%
2	Điểm bài tập	Số bài tập đã làm/số bài tập được giao.	30%
3	Điểm kiểm tra giữa kỳ	- Bài kiểm tra lập trình Python (60 phút)	
4	Điểm thi kết thúc học phần	- Hình thức trắc nghiệm (40 câu/50 phút). - Tham dự đủ 80% tiết lý thuyết. - Dự thi kết thúc học phần.	60%

9.2. Cách tính điểm

- Điểm đánh giá thành phần và điểm thi kết thúc học phần được chấm theo thang điểm 10 (từ 0 đến 10), làm tròn đến một chữ số thập phân.
- Điểm học phần là tổng điểm của tất cả các điểm đánh giá thành phần của học phần nhân với trọng số tương ứng. Điểm học phần tính theo thang điểm 10 làm tròn đến một chữ số thập phân.

phân, sau đó được quy đổi sang điểm chữ và điểm số theo thang điểm 4 theo quy định về công tác đào tạo của Nhà trường.

10. Tài liệu học tập

- [1]. Davy Cielen, Arno D.B. Meysman, Mohamed Ali; Introducing Data Science, Manning Publication Co, 2016. ISBN 978-1-6334-3003-7
- [2]. Jake Vanderplas, Python Data Science Handbook, O'Reilly 2016
- [3] Joel Grus, Data Science from Scratch. O'Reilly Media., 2015. ISBN 978-1-4919-0142-7

11. Hướng dẫn tự học của học phần

Bảng 3. Nội dung chuẩn bị

Tuần	Nội dung	Lý thuyết (tiết)	Bài tập (tiết)	Thực hành (tiết)	Sinh viên cần chuẩn bị
1-2	Chương 1: Mở đầu <ul style="list-style-type: none"> Tổng quan về khoa học dữ liệu Khái niệm về dữ liệu lớn Các ứng dụng của khoa học dữ liệu trong thực tế Tiến trình khoa học dữ liệu Các kiến thức, kỹ năng của một nhà khoa học dữ liệu 	3	0	0	Chapter 1: Data science in a big data world [1] Chapter 2: The data science process [1] Chapter 5: First steps in big data [1]
	Chương 2: Ngôn ngữ lập trình trong khoa học dữ liệu <ul style="list-style-type: none"> Giới thiệu một số ngôn ngữ và công cụ lập trình trong khoa học dữ liệu Python trong khoa học dữ liệu Cài đặt và thiết lập môi trường lập trình Xây dựng ứng dụng đầu tiên với Python 	2	1		Chapter 1: Ipython: Beyond Normal Python [2] Website python: https://www.python.org/
3-6	Chương 3: Lập trình Python căn bản <ul style="list-style-type: none"> Biến, toán tử trong Python Các kiểu dữ liệu cơ bản Cấu trúc điều khiển vòng lặp. 	10	4	0	+ Website w3schools: https://www.w3schools.com/python/default.asp + Youtube: https://www.youtube.com/playlist?list=PLjgjkdf_sn

	<ul style="list-style-type: none"> Hàm, lớp trong Python (def, class) Ngoại lệ và làm việc với các ngoại lệ Làm việc với tập tin Bài tập – Chương 3 				aw8QnlhK5f3DzFDFKD U5f4
7-8	<p>Chương 4: Một số thư viện Python trong khoa học dữ liệu</p> <ul style="list-style-type: none"> Giới thiệu một số Package trong khoa học dữ liệu. Xử lý dữ liệu số với thư viện Numpy Trực quan hóa dữ liệu với thư viện Matplotlib Phân tích dữ liệu với thư viện Pandas Học máy với thư viện Scikit Learn Bài tập – Chương 4 	6	4	0	<p>+ Numpy: https://docs.scipy.org/doc/numpy/user/quickstart.html</p> <p>Chapter 2: Introduction to Numpy [2]</p> <p>+ Matplotlib: https://matplotlib.org/users/index.html</p> <p>- Chapter 9: Data visualization to the end user [1]</p> <p>- Chapter 4: Visualization with Matplotlib [2]</p> <p>+ Scikit Learn: https://scikit-learn.org/stable/</p> <p>- Chapter 3: Machine learning [1]</p> <p>- Chapter 5: Machine learning [2]</p>

Hà Nội, ngày tháng năm 2018

TRƯỞNG KHOA

TRƯỞNG BỘ MÔN

TS Lê Hồng Anh