



**ĐẠI HỌC
BÁCH KHOA HÀ NỘI**
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Nhập môn Khoa học dữ liệu (IT4142)

PGS.TS Thân Quang Khoát & PGS.TS Phạm Văn Hải
Team lecturers

ONE LOVE. ONE FUTURE.

3

Cấu trúc môn học

- Số tuần: 15
 - Lý thuyết: 11-13 tuần
 - Sinh viên trình bày đề án môn học: 02-03 tuần
- Thời gian và địa điểm
 - Theo địa điểm và lịch học của Đại học
- Thời gian gặp sinh viên
 - Hẹn trước qua e-mail
 - Nhà B1



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

4

Môn học này

- Bạn sẽ học cách lấy dữ liệu để
 - Hiểu
 - Xử lý
 - Trích xuất giá trị
 - Trực quan hoá
 - Chia sẻ với người khác
 - Tạo các phán đoán

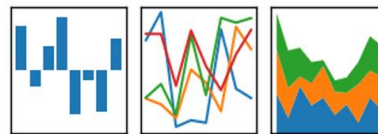


Nội dung môn học

- Lecture 1: Tổng quan về Khoa học dữ liệu
- Lecture 2: Thu thập và tiền xử lý dữ liệu
- Lecture 3: Làm sạch và tích hợp dữ liệu
- Lecture 4: Phân tích và khám phá dữ liệu
- Lecture 5: Trực quan hoá dữ liệu
- Lecture 6: Trực quan hoá dữ liệu đa biến
- Lecture 7: Học máy
- Lecture 8: Phân tích dữ liệu lớn
- Lecture 9: Báo cáo tiến độ bài tập lớn và hướng dẫn
- Lecture 10+11: Phân tích một số kiểu dữ liệu
- Lecture 12: Đánh giá kết quả phân tích



Thư viện hoặc ngôn ngữ



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

7

Đánh giá

- Đồ án môn học (P): Tối đa 10 điểm
 - Mỗi đồ án được thực hiện bởi một nhóm sinh viên
 - Chọn bài toán thực tế muốn giải quyết
 - Chọn một phương pháp phân tích dữ liệu để giải quyết một bài toán thực tế
 - Cài đặt và đánh giá hiệu năng của phương pháp đó dựa trên dữ liệu thực tế
- Thi trắc nghiệm (E): Tối đa 10 điểm
- Điểm học phần (G)
 - $G = 0,4 \times P + 0,6 \times E$



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

8

Đồ án môn học: đề tài

- Tự do đề xuất bài toán thực tế, (các) giải thuật để giải quyết bài toán, và (các) tập dữ liệu được sử dụng
- Đề xuất đề tài phải được **diễn giải cụ thể**
 - **Mô tả bài toán thực tế** sẽ được giải quyết (mục đích, yêu cầu, kịch bản ứng dụng, ...)
 - Xác định rõ **giải thuật** dùng để giải quyết bài toán.
 - Trình bày các thông tin về **đầu vào (input)** và **đầu ra (output)** của hệ thống học máy sẽ được cài đặt, và **cách thức biểu diễn dữ liệu**.
 - Xác định rõ (các) **tập dữ liệu (datasets)** sẽ được sử dụng.
- Phương pháp: Các mô hình AI, ML, thuật toán đề xuất
- Kết quả: Có đánh giá dựa theo mô hình đề xuất hoặc ứng dụng
- Nguồn tham khảo: Các bài báo research papers, thư viện, mã nguồn mở hoặc dự án thực tế



Đồ án môn học: đánh giá

- Công việc đồ án được đánh giá theo các tiêu chí sau:
 - *Mức độ phức tạp / khó khăn của bài toán thực tế được giải quyết*
 - *Chất lượng (sự đúng đắn và phù hợp) của phương pháp được dùng để giải quyết bài toán*
 - *Đánh giá và lựa chọn kỹ lưỡng mô hình*
 - Chất lượng của bài trình bày (presentation) kết quả đồ án
 - Chất lượng của tài liệu báo cáo kết quả đồ án
 - Cài đặt hệ thống thử nghiệm (các chức năng, dễ sử dụng, ...)
 - Video Demo ghi lại các chức năng use case, kết quả
 - Video ghi lại cấu hình, cài đặt và file readme.txt hướng dẫn cài đặt, cấu hình
- Bài trình bày trong khoảng 15 phút, và phù hợp với những gì được nêu trong tài liệu báo cáo
- Nếu sử dụng lại / kế thừa / khai thác các mã nguồn / các gói phần mềm / các công cụ sẵn có, thì **phải nêu rõ ràng và chính xác** trong tài liệu báo cáo (và đề cập trong bài trình bày)



Tài liệu học tập

- Reference books:

- Grus, Joel. *Data science from scratch: first principles with python*. "O'Reilly Media, Inc.", 2015.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition)*. Springer series in statistics, 2009.
- Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.



ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

11



HUST

THANK YOU !



hust.edu.vn



fb.com/dhbkhn

12