

KHAI PHÁ DỮ LIỆU

Bài 0. Giới thiệu môn học

Giáo viên: TS. Trần Mạnh Tuấn

Bộ môn: Hệ thống thông tin

Khoa: Công nghệ thông tin

Email: tmtuan@tlu.edu.vn

Điện thoại: 0983.668.841

Nội dung

1

Giới thiệu môn học

2

Mục tiêu của môn học

3

Phần mềm thực hành

4

Yêu cầu bài tập lớn

Giới thiệu môn học

- Tên môn: Khai phá dữ liệu
- Số tín chỉ: 3 (30 tiết lý thuyết + 15 tiết bài tập)
- Nội dung chính:
 - Tổng quan về ngành khai phá dữ liệu
 - Tiền xử lý dữ liệu
 - Hồi quy dữ liệu
 - Phân lớp dữ liệu
 - Phân cụm dữ liệu
 - Luật kết hợp [1-7]
 - Công nghệ khai phá dữ liệu
 - Ứng dụng của Khai phá dữ liệu

Giới thiệu môn học

- Giảng viên: TS. Trần Mạnh Tuấn, khoa CNTT
TS. Nguyễn Huy Đức, khoa CNTT
ThS. Nguyễn Ngọc Quỳnh Châu, khoa CNTT
- Email: tmtuan@tlu.edu.vn
- Điện thoại: 0983668841

Giới thiệu môn học

■ Tài liệu tham khảo:

- Nguyễn Hà Nam, Nguyễn Trí Thành, Hà Quang Thụy. - Hà Nội ::Đại học Quốc gia Hà Nội,,2016.

■ Đánh giá: ĐQT x **40%** + ĐTCK x **60%**

- Chuyên cần, ý thức: 25%
- Bài tập thực hành: 25%
- Bài kiểm tra: 50%

■ Hình thức đánh giá cuối kỳ: **Vấn đáp BTL**

■ Bài tập lớn

- Nhóm bài tập từ 2–4 sinh viên
- Phân tích thiết kế đầy đủ một đề tài.

Giới thiệu môn học

Khai phá dữ liệu???



Mục tiêu của môn học

- Giới thiệu cho người học tổng quan về các quá trình khám phá tri thức, khai phá dữ liệu, và quá trình tiền xử lý dữ liệu
- Giới thiệu cho người học giá trị lợi ích mà khai phá dữ liệu đóng góp trong các lĩnh vực ứng dụng khác nhau.
- Trình bày các giải thuật và kỹ thuật chính trong giai đoạn tiền xử lý dữ liệu.
- Trình bày các giải thuật và kỹ thuật khai phá dữ liệu chính gồm: hồi qui dữ liệu, phân loại/lớp dữ liệu, gom/phân cụm dữ liệu, và phân tích kết hợp – tương quan (luật kết hợp)
- Tạo khả năng cho người học ứng dụng kỹ thuật KPDL cho các ứng dụng và loại dữ liệu khác nhau

Phần mềm thực hành

- Weka (www.cs.waikato.ac.nz/ml/weka)
- R (www.r-project.org)
- Python
- Tanagra (eric.univ-lyon2.fr/~ricco/tanagra/)
- YALE (rapid-i.com)
- KNIME (www.knime.org)
- Orange (www.ailab.si/orange)
- UCI (<https://archive.ics.uci.edu/ml/index.php>)

Yêu cầu bài tập lớn

- Sinh viên đăng ký bài tập lớn theo nhóm trước ngày 05/02/2021.
- Sinh viên đăng ký tên đề tài từ: 28/02/2021.
- Nộp lần 1: 25/4/2021
- Nộp lần 2: trước khi thi 2 ngày theo lịch thi
- Mỗi bài tập lớn: có ít nhất 2 thuật toán ở 2 lớp bài toán khác nhau.
- Sinh viên xử lý dữ liệu, cài đặt thuật toán, xây dựng ứng dụng.
- Điểm thưởng: không sử dụng các thư viện có sẵn, bài toán có ý nghĩa thực tiễn, dữ liệu không phải chuẩn trên UCI

Trao đổi, câu hỏi?