**Khoa học dữ liệu và học máy**

Thực tế -> Dữ liệu thô -> Xử lý dữ liệu

Khái niệm:

* Cho phép máy móc tự động học từ dữ liệu
* Là 1 bài toán con của trí tuệ nhân tạo
* Dữ liệu thu được là dữ liệu huấn luyện

3 loại tiếp cận chính:

* Học có giám sát: Phân lớp, hồi quy
* Học không có giám sát: Phân cụm dữ liệu, sắp xếp
* Học tăng cường: Liên quan đến điều khiển

Vòng đời: 7 bước

1. Thu thập dữ liệu
2. Chuẩn bị dữ liệu
3. Sắp xếp dữ liệu
4. Phân tích dữ liệu
5. Huấn luyện mô hình
6. Kiểm tra mô hình
7. Đưa vào ứng dụng

Dữ liệu liệt kê, phân loại

1. Import thư viên
2. Import data
3. Kiểm tra các giá trị bị thiếu
4. Kiểm tra các dữ liệu dạng phân loại, phân lớp
5. Biến đổi giá trị các biến thành dải giá trị chung
6. Phân chia dữ liệu thành dữ liệu train và test