1. Dữ liệu là gì?

* Số, chữ, hình ảnh, âm thanh, … của người dùng. Dữ liệu có trước thông tin.

1. Thông tin là gì?

* Là dữ liệu tinh chế (modified)

1. Phân biệt dữ liệu với thông tin (= c1 + c2)
2. Tri thức là gì?

* Hiểu biết về 1 vấn đề

1. Tri thức từ đâu? (nguồn của tri thức)

* Do học được, do bố mẹ truyền cho, do trò chuyện, mạng xã hội

1. Thế nào là tri thức TACIT (tri thức ngầm)

* Không viết ra được, chỉ nằm trong đầu, có từ 1962, do ai? nhà triết học và nhà xã hội học Michael Polany

là loại tri thức mà không thể dễ dàng truyền đạt qua lời nói hoặc viết. Nó thường được hình thành từ kinh nghiệm cá nhân và được tích lũy qua thời gian. Tri thức tacit bao gồm những kỹ năng, trực giác, tinh tế và cảm nhận mà con người có được từ trải nghiệm thực tiễn, thường không thể diễn đạt một cách rõ ràng.

* Vòng xoáy tri thức là gì? (có từ 1998) là một khái niệm trong quản lý tri thức, mô tả quá trình tạo ra, chia sẻ và phát triển tri thức trong tổ chức. Khái niệm này thường được gắn liền với các mô hình như SECI (Socialization, Externalization, Combination, Internalization) của Nonaka và Takeuchi.

1. Hệ thống là gì?

* Tập các thành phần xử lý công việc chung, có đầu ra đầu vào
* VD: bụng của mình.

1. Hệ thống thông tin là gì?

* Hệ thống có i/o là thông tin

1. Phát triển hệ thống là gì?

* Đề xuất thiết kế hệ thống, xây dựng… -> chạy được

1. Thông minh?

* Khả năng sáng tạo, hiểu biết, nhanh nhẹn ,…

1. HTTM là gì?

* Là hệ thống có sử dụng tri thức (trí tuệ nhân tạo)

1. Thuật toán là gì?

* Các bước để giải 1 bài toán

1. Tính chất quan trọng của thuật toán?

* Có i/o, hữu hạn các bước, chính xác, hiệu quả

1. Thể hiện thuật toán ntn?

* Lời văn
* Giả mã
* Sơ đồ khối

1. Học máy là gì?

* Thuộc TTNT, dựa vào những dữ liệu để ktra mô hình
* tập trung vào việc phát triển các thuật toán và mô hình cho phép máy tính học hỏi từ dữ liệu mà không cần phải được lập trình một cách tường minh.
* Máy học mô hình để thực hiện đúng, người sử dụng kết quả ấy

1. Học sâu là gì?

* Là loại học máy sử dụng các mạng nơ-ron sâu (Deep Neural Networks) để mô phỏng cách con người học hỏi từ dữ liệu. Học sâu đặc biệt hiệu quả trong việc xử lý và phân tích các dữ liệu phức tạp như hình ảnh, âm thanh và văn bản.

1. Khác nhau giữa học máy và học sâu?

* Học sâu chính là học máy
* Học sâu là học bằng neron nhân tạo, tự động trích đặc trưng

+ Cấu trúc mô hình

+ Dữ liệu đầu vào

+ Khả năng xử lý dữ liệu phức tạp

+ Thời gian huấn luyện

+ Ứng dụng, yêu cầu tài nguyên

1. Đặc trưng của đối tượng

* Tính chất, cái vốn có của nó

1. Các thành phần vẽ sơ đồ khối

* Mũi tên, mốc (hình tròn), ô thực hiện (hcn), ô kiểm tra (hình thoi)(2 đầu ra, nhiều đầu vào…)

1. Mạng nơ ron nhân tạo ?

* Mạng có nhiều tầng (tầng vào, tầng ẩn, tầng ra)

1. 3 tầng của nơ ron nhân tạo

Mạng nơ-ron nhân tạo là một mô hình tính toán được lấy cảm hứng từ cấu trúc và chức năng của não bộ con người. Nó bao gồm các nơ-ron (hoặc đơn vị xử lý) được kết nối với nhau để xử lý và phân tích thông tin.

Cấu trúc của Mạng Nơ-ron Nhân Tạo:

Tầng vào (Input Layer):

Là tầng đầu tiên của mạng, nơi nhận dữ liệu đầu vào.

Mỗi nơ-ron trong tầng vào tương ứng với một đặc trưng của dữ liệu.

Tầng ẩn (Hidden Layer):

Có thể có một hoặc nhiều tầng ẩn, nơi diễn ra quá trình xử lý chính.

Các nơ-ron trong tầng ẩn nhận đầu vào từ tầng trước và truyền tín hiệu đến tầng tiếp theo.

Tầng ẩn giúp mạng học được các đặc trưng phức tạp từ dữ liệu.

Tầng ra (Output Layer):

Là tầng cuối cùng của mạng, nơi đưa ra kết quả hoặc dự đoán.

Số lượng nơ-ron trong tầng ra phụ thuộc vào bài toán (ví dụ: một nơ-ron cho bài toán hồi quy, nhiều nơ-ron cho bài toán phân loại đa lớp).

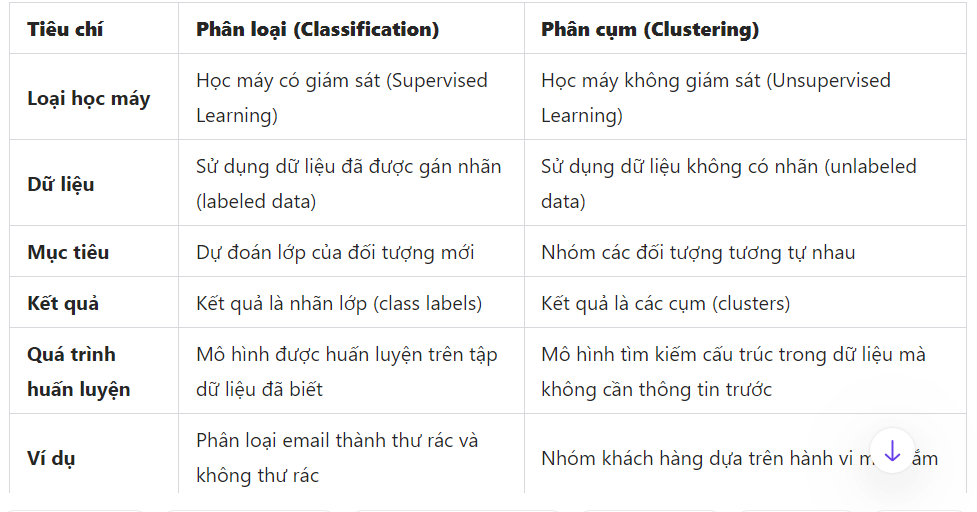
1. Phân loại là gì?

* Chia các đối tượng thành các lớp khác nhau
* nhằm phân nhóm các đối tượng hoặc dữ liệu vào các danh mục hoặc lớp khác nhau dựa trên các đặc trưng của chúng.

1. Phân cụm là gì?

* Chia các phần tử thành các cụm khác nhau
* dùng để nhóm các đối tượng hoặc dữ liệu thành các cụm (clusters) sao cho các đối tượng trong cùng một cụm có đặc điểm tương tự nhau hơn so với các đối tượng ở các cụm khác.

1. Khác nhau giữa phân loại phân cụm?

* Phân loại phải gán nhãn trước.
* Phân cụm tự động gán nhãn
* 

1. Bộ dữ liệu ban đầu để học máy được tách ra?

* (huấn luyện/thử nghiệm/kiểm chứng)
* Bộ dữ liệu huấn luyện (Training Set):
* Là phần lớn nhất của bộ dữ liệu, được sử dụng để huấn luyện mô hình.
* Mô hình học từ bộ dữ liệu này để tìm ra các mẫu và mối quan hệ trong dữ liệu.
* Bộ dữ liệu kiểm tra (Test Set):
* Là phần dữ liệu được tách ra để đánh giá hiệu suất của mô hình sau khi đã được huấn luyện.
* Bộ dữ liệu này không được sử dụng trong quá trình huấn luyện, nhằm đảm bảo rằng mô hình có thể tổng quát hóa tốt cho dữ liệu chưa thấy.
* Bộ dữ liệu xác thực (Validation Set) (nếu cần):
* Được sử dụng để điều chỉnh các tham số của mô hình và chọn mô hình tốt nhất.
* Giúp kiểm tra hiệu suất của mô hình trong quá trình huấn luyện mà không cần phải sử dụng bộ kiểm tra.
* Kiểm tra tính chính xác, tối ưu hóa cho mô hình

1. Thế nào là gán nhãn cho 1 đối tượng?

Gán nhãn cho một đối tượng (labeling) là quá trình xác định và chỉ định một nhãn (label) cụ thể cho một đối tượng hoặc mẫu trong tập dữ liệu

1. Thế nào là mô hình học máy?

* Là 1 mô hình mẫu (có những mô hình nào)
* Mô hình học máy là một hệ thống hoặc thuật toán được thiết kế để học từ dữ liệu và đưa ra dự đoán hoặc quyết định mà không cần được lập trình cụ thể cho từng nhiệm vụ.
* Học có giám sát (Supervised Learning):
* Mô hình hồi quy (Regression Models): Dùng để dự đoán giá trị liên tục.
* Ví dụ: Hồi quy tuyến tính, hồi quy logistic.
* Mô hình phân loại (Classification Models): Dùng để phân loại dữ liệu thành các nhóm khác nhau.
* Ví dụ: Cây quyết định (Decision Trees), hồi quy logistic, mạng nơ-ron (Neural Networks), SVM (Support Vector Machines).
* Học không giám sát (Unsupervised Learning):
* Phân cụm (Clustering): Nhóm các đối tượng tương tự nhau.
* Ví dụ: K-means, phân cụm phân cấp (Hierarchical Clustering).
* Giảm chiều dữ liệu (Dimensionality Reduction): Giúp giảm số lượng biến trong dữ liệu.
* Ví dụ: PCA (Principal Component Analysis), t-SNE.

1. THế nào là thống kê?

* Thống kê dựa trên các quan sát để chứng minh giả thuyết
* Thống kê là một lĩnh vực khoa học liên quan đến việc thu thập, phân tích, diễn giải, trình bày và tổ chức dữ liệu.

1. Mấy loại thống kê?

* Thống kê mô tả và thống kê suy luận
* Thống kê mô tả: Dùng để tóm tắt và mô tả dữ liệu từ một mẫu, bao gồm các biện pháp như giá trị trung bình, độ lệch chuẩn, và độ lệch trung bình.
* Thống kê suy luận: Dùng để rút ra kết luận từ dữ liệu mẫu và áp dụng cho toàn bộ dân số, bao gồm các phương pháp như phân tích hồi quy và kiểm định giả thuyết1

1. Trung vị là gì? Hãy cho thí dụ minh hoạ

* Trung vị (Median) là giá trị nằm ở vị trí trung tâm của một dãy số đã được sắp xếp theo thứ tự từ thấp đến cao hoặc từ cao đến thấp. Trung vị giúp mô tả dữ liệu mà không bị ảnh hưởng bởi các giá trị ngoại lai (quá lớn hoặc quá nhỏ).