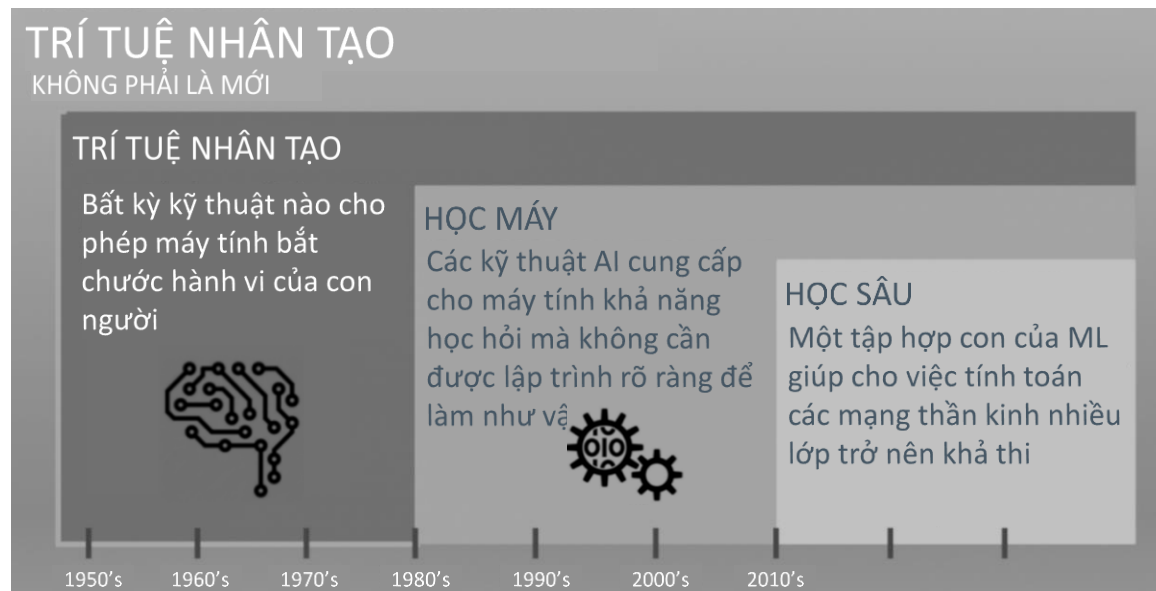


I Xây dựng một hệ thống thông minh chuyển đổi dữ liệu thành tri thức

- ▶ Có rất nhiều dữ liệu có cấu trúc và phi cấu trúc trong thời đại công nghệ hiện đại.
- ▶ Học máy xuất hiện vào cuối thế kỷ 20 như một lĩnh vực phụ của Trí tuệ nhân tạo (AI) liên quan đến các thuật toán tự học trích xuất và dự đoán kiến thức từ dữ liệu.
- ▶ Con người phân tích thủ công một lượng lớn dữ liệu để tạo ra các quy tắc và tạo mô hình.
- ▶ Học máy có thể dần dần cải thiện các mô hình dự đoán và hiệu suất ra quyết định dựa trên dữ liệu bằng cách trích xuất kiến thức hiệu quả hơn từ dữ liệu.



Phân loại Trí tuệ nhân tạo



- | “Một lĩnh vực nghiên cứu phát triển các thuật toán cho phép máy học từ dữ liệu và thực hiện các hành động không được chỉ định rõ ràng bằng mã” - Arthur Samuel, 1959
- | “Một chương trình máy tính được cho là học hỏi từ trải nghiệm E đối với một số loại nhiệm vụ T và thước đo hiệu suất P nếu hiệu suất của nó ở các nhiệm vụ trong T, được đo bằng P, được cải thiện nhờ trải nghiệm E.” - Tom Mitchell, 1977

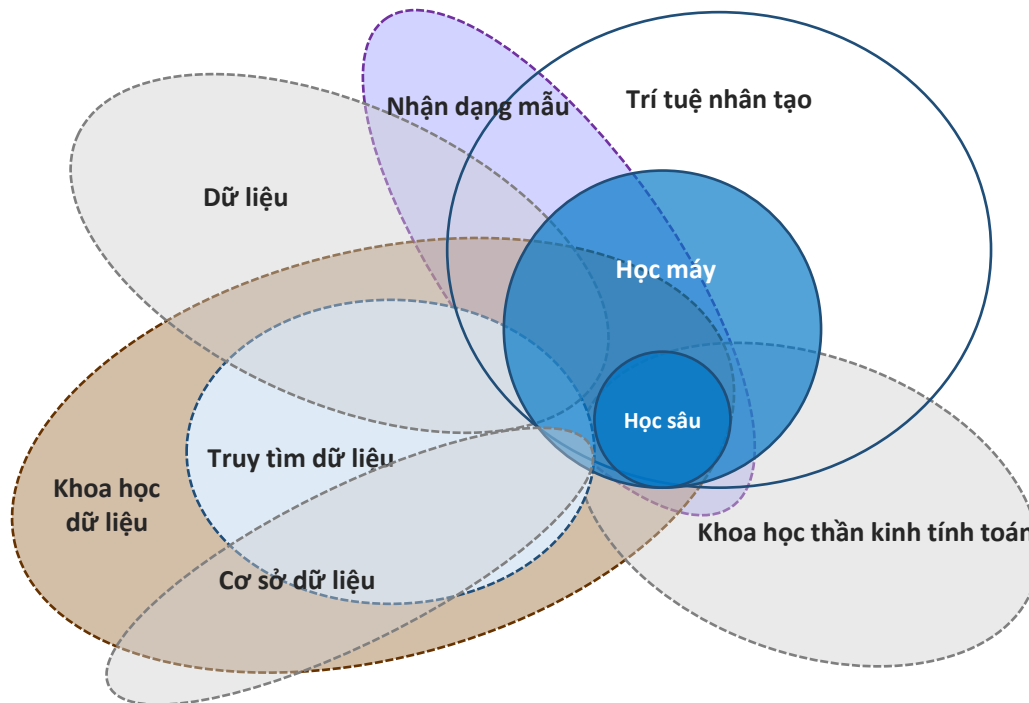


- | Nhiệm vụ T là phân loại chó và mèo, và hiệu suất P thể hiện thước đo phân loại chó và mèo. E có thể nói là “đang học” nếu hiệu suất phân loại chó, mèo nâng cao dần qua kinh nghiệm, hoặc dữ liệu (10.000 ảnh).

Các ngành liên quan đến học máy

I Các nghiên cứu liên quan đến học máy

- ▶ Học máy là một lĩnh vực liên ngành kết hợp nền tảng học thuật và thành tựu trong các lĩnh vực khác nhau như xác suất và thống kê, khoa học máy tính, lý thuyết cơ sở dữ liệu, khoa học nhận thức, khoa học thần kinh và nhận dạng mẫu, thay vì chỉ giới hạn ở công nghệ phương pháp luận trong bất kỳ lĩnh vực nào.



Mối quan hệ giữa học máy và các ngành khác nhau

I Học máy và Thống kê số liệu

- ▶ Theo truyền thống, **thống kê** được coi là nền tảng lý thuyết cung cấp một phương pháp khoa học và có hệ thống để chuyển đổi dữ liệu thành thông tin. Các lĩnh vực được nhấn mạnh đặc biệt trong thống kê là suy luận và xác minh, và các phương pháp và lý thuyết khác nhau đã được thiết lập để **giải thích mức độ phù hợp của dữ liệu đã cho với giả thuyết và lý thuyết của nhà nghiên cứu** (hoặc mức độ chính xác của các giá trị thu được từ các quan sát ước tính các tham số dân số thực tế).
- ▶ Trong khi đó, **học máy** chủ yếu được sử dụng để giải quyết các nhiệm vụ khó thiết kế hoặc khó lập trình các thuật toán rõ ràng. Hầu hết các thuật toán học máy trước tiên được sử dụng để định lượng các **mối quan hệ phức tạp bằng cách xác định tính năng của các cơ chế tiềm năng do dữ liệu tạo ra**, sau đó đưa ra dự đoán về dữ liệu mới bằng cách sử dụng mẫu đã xác định này.
- ▶ Thoạt nhìn, các cách tiếp cận của thống kê và học máy có vẻ đối lập nhau, nhưng bên cạnh sự nhấn mạnh vào **một số khác biệt về khía cạnh hoặc quan điểm**, các phương pháp hình thành nền tảng của mỗi ngành học là rất giống nhau. Trên thực tế, có thể nói rằng nhiều phương pháp học máy dựa trên việc học thống kê dựa trên số liệu thống kê.

