Machine learning

1. Giới thiệu về học máy (<https://www.youtube.com/watch?v=NVsw-JrXv9I>)

(Lập trình AI bằng Tensor Flow)

(<https://www.youtube.com/watch?v=13R2ytXz7j0>)

Machine Learning chỉ hành động máy tính tự học để cải thiện một nhiệm vụ mà nó đang thực hiện. Hiệu suất của máy tính khi thực hiện một nhiệm vụ sẽ trở nên tốt hơn sau khi hoàn thành nhiệm vụ đó nhiều lần. Hay nói cách khác khả năng cơ bản nhất của Machine Learning là dùng thuật toán để phân tích những thông tin có sẵn, học hỏi từ nó rồi đưa ra quyết định hoặc dự đoán gì đó có liên quan.

Ví dụ: Công cụ tìm kiếm của google. Khi bạn đang tìm kiếm gì đó trên google nó sẽ trả lại rất nhiều kết quả tìm kiếm. Nếu bạn bỏ nhiều thời gian để xem những kết quả trả về hoặc click vào đường link nào để đọc tìm thì gg sẽ ghi nhận người đó bỏ nhiều thời gian để xem thông tin đó có nghĩa là thông tin này hữu ích và thông tin phù hợp với người này. Nếu bạn chỉ xem lướt qua thì gg sẽ hiểu thông tin trả về không phù hợp. -> Machine Learning sẽ phỏng đoán được những điều bạn thích và không thích và sẽ tự điều chỉnh những kết quả phù hợp nhất dành cho bạn.

(<https://vi.wikipedia.org/wiki/H%E1%BB%8Dc_m%C3%A1y>)

Machine Learning là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép hệ thống “học “ tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Các thuật toán máy xây dựng mô hình dựa trên dữ liệu mẫu, được gọi là dữ liệu huấn luyện (**traning set**), để đưa ra dự đoán hoặc quyết định mà không cần được lập trình chi tiết về việc đưa ra dự đoán hoặc quyết định này.

Ví dụ như các máy học có thể học cách phân loại thư điện tử xem có phải thư spam hay không và tự động sắp xếp thư vào mục tương ứng.

Học máy có hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chuẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động robot

Thuật toán Random Forest là một trong các thuật toán học máy ra đời muộn hơn các thuật toán học máy khác, chính vì vậy thuật toán Random Forest kế thừa được ưu điểm các thuật toán khác đồng thời khắc phục được hạn chế về mặt số lượng dữ liệu cũng như độ phức tạp của dữ liệu.

(https://tek4.vn/khoa-hoc/machine-learning-co-ban/training-set-va-testing-set)

*Training set – là tập dữ liệu được sử dụng để huấn luyện mô hình. Các thuật toán học máy sẽ học các mô hình từ tập huấn luyện này. Việc học sẽ khác nhau tùy thuộc vào thuật toán và mô hình sử dụng.*

(<https://www.mcivietnam.com/blog-detail/kham-pha-training-set-testing-set-va-validation-set-trong-phan-chia-du-lieu/>)

1. Training set – là một tập dữ liệu dùng để huấn luyện cho mô hình của thuật tonas machine learning. Đây thường là một tập dữ liệu có kích thước lớn, được dùng để training trong quá trình huấn luyện máy học.
2. Testing set – là tập dữ liệu dùng để test sau khi máy tính đã được học xong. Một mô hình máy học sau khi được huấn luyện, sẽ cần phải được kiểm chứng xem có đạt hiệu quả không. Mô hình học máy cũng vậy, sau mỗi quá trình huấn luyện, các mô hình này sẽ được kiểm chứng độ chính xác, nếu đáp ứng được yêu cầu là đạt, không thì phải xem xét lại. Và để kiểm nghiệm được độ chính xác của mô hình này, người ta dùng testing set. Khác với training set, testing set chỉ gồm các giá trị input (TD, GC, YKNTK, TI, RRCN) mà không có giá trị output (YDM) 🡪 Testing set là tập các giá trị input và được dùng để kiểm thử độ chính xác của những mô hình máy học sau khi huấn luyện.
3. Validation Set – là tập các giá trị input đi kèm các giá trị output và được dùng để kiểm thử độ chính xác của mô hình máy học trong quá trình huấn luyện.
4. (<https://vi.wikipedia.org/wiki/Ch%E1%BA%A9n_%C4%91o%C3%A1n>) Chuẩn đoán là xác định tính chất và nguyên nhân của sự vật hiện tượng. Chuẩn đoán được sử dụng nhiều trong nhiều ngành nhiều biến thể khác nhau bằng cách sử dụng biện luận, phân tích và kinh nghiệm để phát hiện ra nguyên nhân kết quả. Trong kĩ thuật hệ thống và khoa học máy tính, chuẩn đoán thường được sử dụng để xác định nguyên nhân của triệu chứng, cách giảm nhẹ hay cách giải quyết vấn đề.