**Báo cáo task2**

Học máy về bản chất là máy học, là ngành khoa học giúp cho máy tính có được tri giác, được cảm nhận như con người về một vấn đề nào đó thông qua việc “học”.Học ở đây là việc nghiên cứu các tình huống đã xảy ra từ trước nhằm giúp máy tính xây dựng ra bộ “cảm quan” của mình từ đó phân tích các sự vật sự việc dưới cái nhìn cảm quan riêng của mình. Càng nhiều dữ liệu chuẩn , sạch và đầy đủ được đưa vào thì cảm quan của máy sẽ càng chính xác

Học máy chính hợp thành từ 3 ngành khoa học cơ bản: khoa học máy tính, thống kê và toán học

Học máy có thể được ứng dụng trong việc nhận dạng các mẫu(pattern recognition), khai phá dữ liệu(data mining), phân tích dữ liệu(data science), dự đoán (redictive analytics), ............

Phân loại học máy:  
 -Học có giám sát là thuật toán dự đoán đầu ra (outcome) của một dữ liệu mới (new input) dựa trên các cặp (input, outcome) đã biết từ trước. Cặp dữ liệu này còn được gọi là (data, label), tức (dữ liệu, nhãn). Gồm

+ Phân loại: Chia nhãn của dữ liệu đầu vào thành một số nhóm hữu hạn

+ Hồi quy: Cũng chia nhóm nhưng có giá trị cụ thể

-Học không giám sát, không biết được dữ liệu đầu ra. Thuật toán unsupervised learning sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm (clustering) hoặc giảm số chiều của dữ liệu (dimension reduction) để thuận tiện trong việc lưu trữ và tính toán.

+ Phân cụm: Một bài toán phân nhóm toàn bộ dữ liệu X thành các nhóm nhỏ dựa trên sự liên quan giữa các dữ liệu trong mỗi nhóm.

+Association: Khai phá ra các luật dựa trên thằng dữ liệu cho trước

-Học bán giám sát: có một lượng lớn dữ liệu X nhưng chỉ một phần trong chúng được gán nhãn được gọi là Semi-Supervised Learning. Những bài toán thuộc nhóm này nằm giữa hai nhóm được nêu bên trên. Vd: phân tích ảnh

-Học củng cố: toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất (maximizing the performance).

**Các bước của một bài toán học máy bao gồm:**

1. Định nghĩa vấn đề:

+Mô tả vấn đề :

-Mô tả không chính thức

-Mô tả chính thức

-Giả thuyết

+Các dữ liệu được cung cấp

-Ràng buộc áp đặt về dữ liệu

- Định dạng các thuộc tính

+Sự thúc đẩy

-Lợi ích

- Sử dụng

+Giải pháp thủ công

1. Phân tích dữ liệu

+Tóm tắt dữ liệu

-Cấu trúc dữ liệu

-Phân bố dữ liệu

+Hiển thị hóa dữ liệu

-Biểu đồ thuộc tính

- Các thuộc tính ghép cặp phân tán

1. Chuẩn bị dữ liệu

+Lựa chọn dữ liệu

+Tiền xử lý dữ liệu

-Định dạng

-Làm sạch

-Lấy mẫu

+Cải tạo dữ liệu

-Phân tích(Xáo trộn?)

-Gộp dữ liệu

1. Đánh giá thuật toán

+Khai thác thử nghiệm và các tùy chọn

+Tìm và lựa chọn các thuật toán

+Giải thích và báo cáo kết quả

1. Cải thiện kết quả

+Điều chỉnh thuật toán

+Đồng bộ

* Đóng gói,
* Cải thiện
* Làm hòa hợp

+Kĩ thuật cực đoan

1. Trình bày kết quả

+Trình bày kết quả

-Ngữ cảnh

-Vấn đề

-Giải pháp

-Phát hiện

-Giới hạn

-Kết luận

+Phân tán hoạt động