1. Introduction

* Bệnh tim để chỉ một loạt các tình trạng có ảnh hưởng đến trái tim – từ các nhiễm trùng đến các khuyết tật di truyền và các bệnh mạch máu. Hầu hết bệnh tim có thể được ngăn ngừa bằng cách lựa chọn lối sống lành mạnh, nhưng hiện tại, nó vẫn là mối đe dọa sức khỏe số một trên thế giới. Dưới đây là các con số thống kế cảnh báo liên quan đến bệnh tim:
  + 200 000 người ở Việt Nam chết vì bệnh tim mỗi năm – chiếm ¼ ca tử vong. Mỗi năm, cứ 3 người Việt có 1 người bị nhồi máu cơ tim.
  + 24,3% và 24,1% tương ứng là tỉ lệ các ca tử vong của người da trắng và người Mỹ gốc Phi.
  + Hơn 80% các ca tử vong trên thế giới do bệnh tim xảy ra ở các nước thu nhập thấp và thu nhập trung bình. Năm quốc gia hang đầu có tỷ lệ tử vong cao nhất do bệnh tim là:
    - Nga
    - Bulgaria
    - Hungary
    - Argentina
* Các yếu tố làm tăng gấp đôi nguy cơ mắc bệnh tim:
  + Huyết áp cao: 75% người bị suy tim mãn tính có huyết áp cao.
  + Cholesterol cao: Những người có cholesterol cao có khả năng phát triển bệnh tim gấp 2 lần so với những người có mức cholesterol bình thường.
  + Bệnh tiểu đường: Người bị tiểu đường có khả năng phát triển bệnh tim gấp 2 lần so với những người không có nó.
  + Trầm cảm: Những người bị trầm cảm có hơn 25–40% khả năng chết vì bệnh tim so với những người không bị trầm cảm.
  + Béo phì: Bệnh động mạch vành hiện diện nhiều hơn gấp 10 lần ở những người bị béo phì. Béo phì là khi chỉ số khối cơ thể (BMI) bằng 30 hoặc cao hơn. 20% trẻ em trên 5 tuổi và 35% người lớn được xem là béo phì.
* Nhận thấy tính cấp bách và mức độ quan trọng của các loại bênh về tim, nhóm chúng em nhận thấy đề tài “Áp dụng Machine Learning vào để chấn đoán bệnh tim” là một đề tài thú vị, giúp cho chúng em có thêm kiến thức về y khoa cũng như được nghiên cứu 1 đồ án có ý nghĩa cho xã hội.
* Những điều chúng em sẽ đạt được khi hoàn thành đồ án này:
  + Kiến thức cơ bản về Machine Learning.
  + Kiến thức y khoa cơ bản về tim.
  + Khả năng làm việc nhóm.
  + Khả năng chịu đựng áp lực công việc.

1. Goal

* Khi đồ án được hoàn thành, sản phẩm này có thể:
  + Dựa vào ảnh điện tâm đồ có thể giúp bác sĩ chẩn đoán bệnh với tỉ lệ chính xác lên đến 73%.
  + Khả năng tự học thêm nhiều loại bênh mới về tim qua các chuyên gia đầu ngành.
  + Khả năng cảnh báo khi tim có dấu hiệu bất thường.

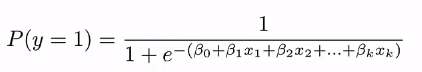
1. Objectives
   * Tìm hiểu về bệnh tim mạch
   * Tìm hiểu về máy điện tâm đồ, cách sử dụng máy và hiểu các thông số cơ bản của máy.
   * Tìm hiểu về thuật toán Logistic Regression và Random Forest.
   * Biết một số ngôn ngữ như C#, Python để lập trình ứng dụng.
   * Nắm kiến thức của môn đại số tuyến tính và xác suất ứng dụng để xử lí tình huống.
2. Theory
   1. Kiến thức vận dụng
   * Thuật toán Logistic Regression
   * Là một mô hình hồi quy nhằm dự đoán giá trị đầu ra *rời rạc* (*discrete target variable*) *y* ứng với một véc-tơ đầu vào x tương đương với chuyện phân loại các đầu vào x vào các nhóm *y* tương ứng.
   * Thuật toán Random Forest
   * Là một thuật toán có giám sát, hoạt động bằng cách đánh giá nhiều Cây quyết định ngẫu nhiên, và lấy ra kết quả được đánh giá tốt nhất trong số kết quả trả về.
   1. Triển khai
   * Thuật toán Logistic Regression
     + Sử dụng biến phụ thuộc để phân loại mắc bệnh và không mắc bệnh.
     + Giả sử gọi biến phụ thuộc y là mắc bệnh ( y=1: mắc bệnh, y=0: không mắc bệnh ).
     + Dự báo P(y=1)
     + P(y=0) = 1 – P(y=0)
     + Các biến độc lập khác x1,x2,…..xk, ví dụ:

.Tuổi tác

.Giới tính

.Đặc điểm,…..

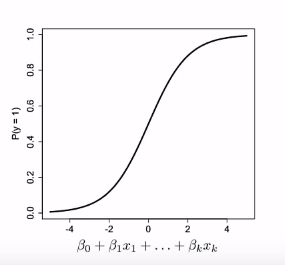
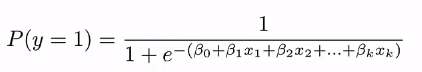
* + - Sử dụng hàm Logistic:



---> Đây là phép biến đổi để đưa ra giá trị dự báo trong khoảng (0,1)

* + - Hàm Logistic

Lí do chọn: cùng một khoảng giá trị xét nhưng mô hình logistic cho kết quả tốt hơn mô hình probit ( probit cho kết quả nằm ngoài khoảng giá trị xét (0,1) ).

. Hàm số cho ở trên luôn cho giá trị dự báo luôn nằm trong khoảng (0,1).

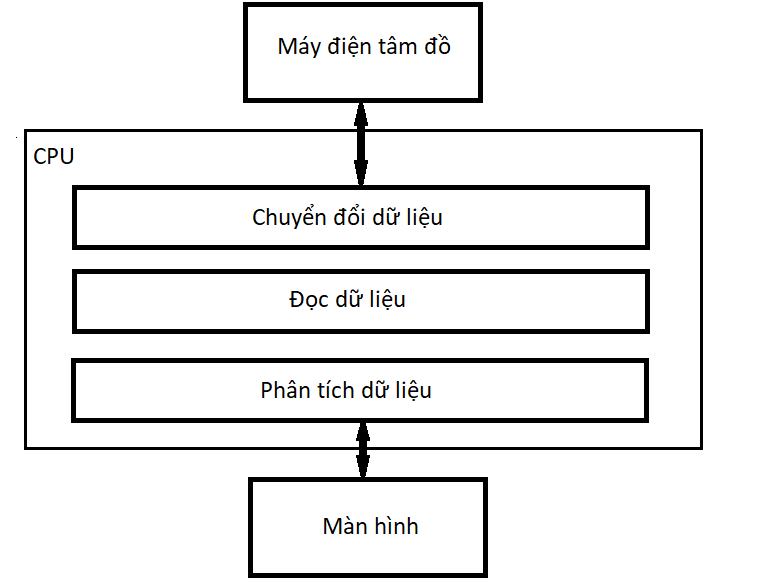
. Các tham số của mô hình được ước lượng sao cho:

Dự báo xác suất lớn cho trường hợp mắc bệnh

Dự báo xác suất nhỏ cho trường hợp không mắc bệnh.

* Thuật toán Random Forest
  + *Chọn ngẫu nhiên* ***“k”*** *features từ tập* ***“m”*** *features.*
  + *Để ý* ***k << m***
  + *từ tập* ***“k”*** *features, tính toán ra node* ***“d”*** *là tốt nhất cho Node phân loại.*
  + *Chia các node con theo node tốt nhất vừa tìm được*
  + *Lặp lại bước* ***1-3*** *cho đến khi đạt đến* ***k*** *node*
  + *Lặp lại bước* ***1-4*** *để tạo ra* ***“n”*** *cây*

1. UML
   1. Block Diagram



* 1. Flowchart

