**TRƯỜNG ĐẠI HỌC THỦ DẦU MỘT**

**VIỆN KỸ THUẬT CÔNG NGHỆ**



ĐỒ ÁN MÔN HỌC

**HỌC MÁY**

Đề tài

XÂY DỰNG CÂY QUYẾT ĐỊNH CHO TRÒ CHƠI TIC TAC TOE

GVHD: ThS. Bùi Thanh Khiết Nhóm thực hiện: 09

|  |  |  |
| --- | --- | --- |
| **STT** | **Họ và tên** | **Lớp** |
| 1 | Mai Văn Chánh | D18PM04 |
| 2 | Hồ Diên Công | D18PM04 |
| 3 | Hồ Diên Thành | D18PM04 |
| 4 | Bùi Thành Được | D18PM01 |
| 5 | Bùi Hoàng Xuân | D18PM04 |

Tháng 07 năm 2021

**VIỆN KỸ THUẬT – CÔNG NGHỆ**

**NHẬN XÉT VÀ CHẤM ĐIỂM CỦA GIẢNG VIÊN**

Họ và tên giảng viên: **Bùi Thanh Khiết**

Tên đề tài: xây dựng cây quyết định cho trò chơi tic tac toe

Nội dung nhận xét:

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

...........................................................................................................................................

**Điểm:**

Bằng số: .................................................................

Bằng chữ:................................................................

**GIẢNG VIÊN CHẤM**

*(Ký, ghi rõ họ tên)*

**BÙI THANH KHIẾT**

# MỤC LỤC

Contents

[MỤC LỤC 2](#_Toc80448793)

[DANH SÁCH HÌNH ẢNH 4](#_Toc80448794)

[LỜI MỞ ĐẦU 1](#_Toc80448795)

[CHƯƠNG 1: GIỚI THIỆU 2](#_Toc80448796)

[1. Tổng quan về đề tài 2](#_Toc80448797)

[2. Mô tả bài toán 2](#_Toc80448798)

[3. Các bước thực hiện cho bài toán 3](#_Toc80448799)

[4. Công cụ thực hiện - Visual Studio 3](#_Toc80448800)

[CHƯƠNG 2: CƠ SỞ LÍ THUYẾT 5](#_Toc80448801)

[1.Tập dữ liệu đầu vào 5](#_Toc80448802)

[1.1 Cấu trúc dữ liệu 5](#_Toc80448803)

[1.2 Chuẩn hóa dữ liệu 6](#_Toc80448804)

[1.3 Dữ liệu mẫu 6](#_Toc80448805)

[2. Cây quyết định 8](#_Toc80448806)

[2.1 Giới thiệu 8](#_Toc80448807)

[2.2 Các kiểu cây quyết định 9](#_Toc80448808)

[2.3 Ưu điểm của cây quyết định 9](#_Toc80448809)

[3.3 Nhược điểm của cây quyết định 10](#_Toc80448810)

[3. Thuật toán ID3 11](#_Toc80448811)

[3.1 Giới thiệu ID3 11](#_Toc80448812)

[3.2 Entropy 11](#_Toc80448813)

[3.3 Information Gain 13](#_Toc80448814)

[CHƯƠNG 3: QUY TRÌNH XÂY DỰNG VÀ KẾT QUẢ 14](#_Toc80448815)

[1. Xây dựng phần mềm 14](#_Toc80448816)

[1.1 khai báo thư viện 14](#_Toc80448817)

[1.2 đọc dữ liệu từ csv 14](#_Toc80448818)

[1.3 chia thuộc tính và kết quả 14](#_Toc80448819)

[1.4 chia data thành training và testing 14](#_Toc80448820)

[1.5 tạo đối tượng cây quyết định loại entropy 14](#_Toc80448821)

[1.6 Huấn luyện 14](#_Toc80448822)

[1.7 dự đoán 14](#_Toc80448823)

[1.8 tính độ chính xác 15](#_Toc80448824)

[1.9 vẽ cây 15](#_Toc80448825)

[2. Kết quả thu được 16](#_Toc80448826)

[2.1 Giao diện nhập dữ liệu đầu vào 16](#_Toc80448827)

[2.2 Giao diện kết quả 17](#_Toc80448828)

[2.3 Cây quyết định của tập dữ liệu mẫu 18](#_Toc80448829)

[CHƯƠNG 4: KẾT LUẬN 19](#_Toc80448830)

[1. Kết quả đạt được 19](#_Toc80448831)

[2. Kết quả chưa đạt được và khó khăn 19](#_Toc80448832)

[3. Hướng phát triển 19](#_Toc80448833)

[TÀI LIỆU THAM KHẢO 20](#_Toc80448834)

# DANH SÁCH HÌNH ẢNH

[Hình 1 : Đề tài mô phỏng 2](#_Toc80448103)

[Hình 2:Dữ liệu ban đầu chưa được chuẩn hóa 6](#_Toc80448104)

[Hình 3:Dữ liệu sau khi được chuẩn hóa 7](#_Toc80448105)

[Hình 4 : Entropy 12](#_Toc80448106)

[Hình 5:Giao diện nhập dữ liệu đầu vào 16](#_Toc80448107)

[Hình 6: Giao diện kết quả 17](#_Toc80448108)

[Hình 7: Cây quyết định của tập dữ liệu mẫu 18](file:///C:\Users\AnThony%20Tony\Desktop\Nhom2_HocMay-đã%20chuyển%20đổi.docx#_Toc80448109)

# LỜI MỞ ĐẦU

Hiện nay, cùng với sự phát triển của khoa học kĩ thuật, nhu cầu giải trí ngày càng nhiều các trò chơi ngày càng phổ biến, trong đó Tic-tac-toe là một trò chơi phổ biến dùng viết trên bàn cờ giấy có chín ô, 3x3. Hai người chơi, người dùng ký hiệu O, người kia dùng ký hiệu X, lần lượt điền ký hiệu của mình vào các ô. Người thắng là người thể tạo được đầu tiên một dãy ty ký hiệu của mình, ngang dọc hay chéo đều được.

Thông qua việc xử lý dữ liệu ta có thể lấy được kết quả X thằng hoặc thua. Từ đó ta có thể xây dựng một cây quyết định cho trò chơi với hệ thống tự chơi với con người theo các cấp độ khác nhau, có thể xử lý các tình huống nước đi của đối thủ và giành chiến thắng hoặc thất bại

Chính vì vậy nên nhóm chúng em đã quyết định chọn đề tài

“xây dựng cây quyết định cho trò chơi tic tac toe ”.

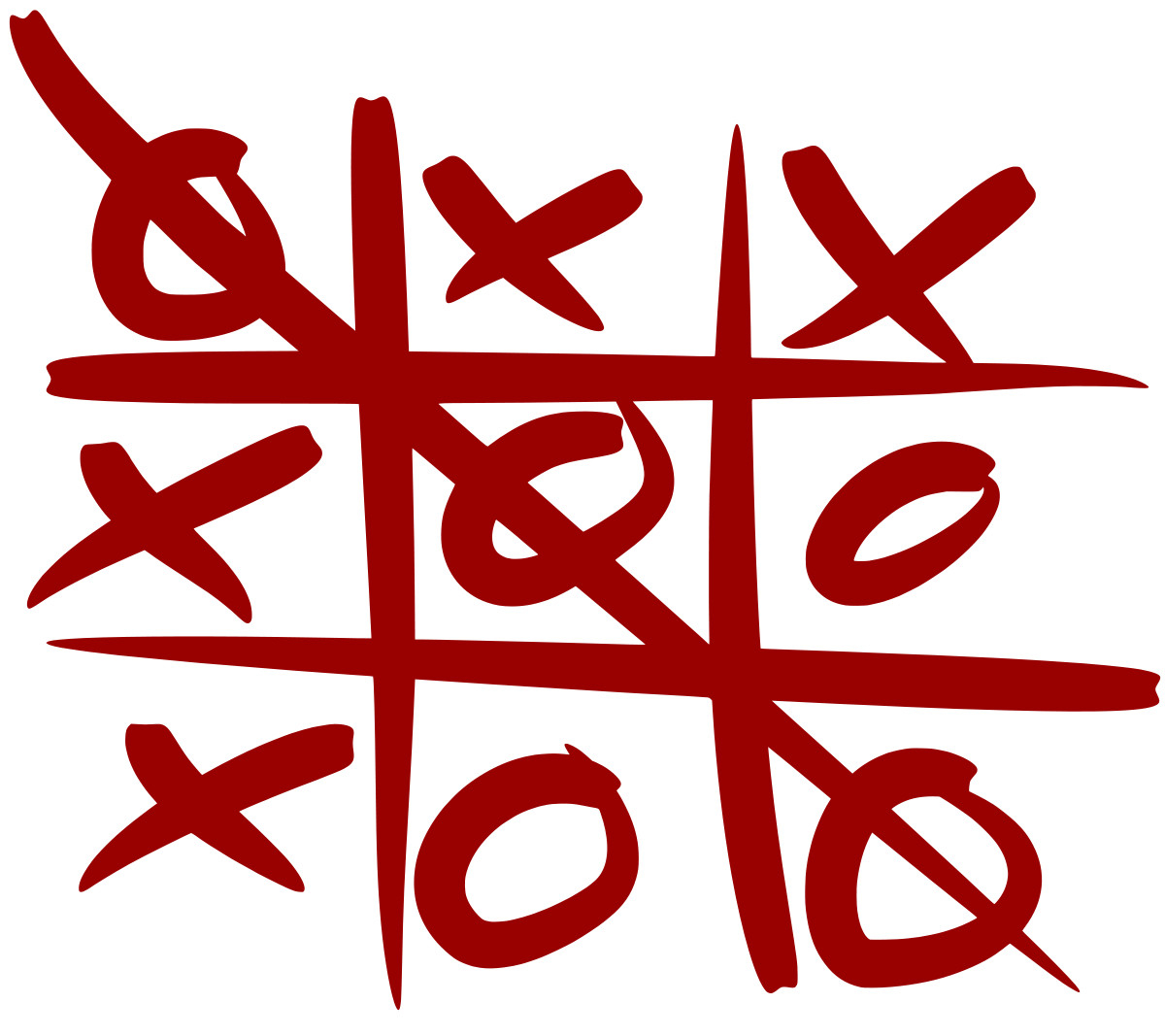
Trong quá trình học tập và được sự giúp đỡ tận tình của thầy nhưng do một số khó khăn trong học tập và ảnh hưởng từ dịch bệnh nên việc làm báo cáo của chúng em trong quá trình thực diện dự án sẽ phát sinh sai sót. Chúng em mong nhận được sự góp ý từ thầy để chúng em có thể phát triển và hoàn thiện hơn bài báo cái của nhóm .

Chúng em xin chân thành cám ơn!

# CHƯƠNG 1: GIỚI THIỆU

## Tổng quan về đề tài

xây dựng trò chơi tic tac toe là phần mềm có thể tự triển khai các nước đi của mình thông qua dữ liệu data thu thập được , gải quyết vấn đề lever người chơi.



Hình 1 : Đề tài mô phỏng

## Mô tả bài toán

Phần mềm sử dụng tập dữ liệu đầu vào là tập các giá trị của các thuộc tính được thu thập lại từ các nước đi của đối thủ từ đó xử lý các thông tin và giải quyết bài toán. Ứng với mỗi nước đi sẽ có nhiều giá trị tương ứng. Sau khi được huấn luyện thì kết quả sẽ là cây quyết định

## Các bước thực hiện cho bài toán

Bước 1: Tìm kiếm tập dữ liệu đầu vào Bước 2: Chuẩn hóa tập dữ liệu đầu vào Bước 3: Đọc dữ liệu đầu vào

Bước 4: tạo đối tượng cây quyết định loại entropy

Bước 5: Huấn luyện phần mềm bằng tập dữ liệu đầu vào Bước 6: Tính độ chính xác

Bước 7: Vẽ cây tìm kiếm

## Công cụ thực hiện - Visual Studio

Visual Studio là một môi trường phát triển tích hợp (IDE) từ Microsoft. Microsoft Visual Studio còn được gọi là “Trình soạn thảo mã nhiều người sử dụng nhất thế giới “, được dùng để lập trình C++ và C# là chính. Nó được sử dụng để phát triển chương trình máy tính cho Microsoft Windows, cũng như các trang web, các ứng dụng web và các dịch vụ web. Visual Studio sử dụng nền tảng phát triển phần mềm của Microsoft như Windows API, Windows Forms, Windows Presentation Foundation, Windows Store và Microsoft Silverlight. Nó có thể sản xuất cả hai ngôn ngữ máy và mã số quản lý.

Visual Studio bao gồm một trình soạn thảo mã hỗ trợ IntelliSense cũng như cải tiến mã nguồn. Trình gỡ lỗi tích hợp hoạt động cả về trình gỡ lỗi mức độ mã nguồn và gỡ lỗi mức độ máy. Công cụ tích hợp khác bao gồm một mẫu thiết kế các hình thức xây dựng giao diện ứng dụng, thiết kế web, thiết kế lớp và thiết kế giản đồ cơ sở dữ liệu. Nó chấp nhận các plug-in nâng cao các chức năng ở hầu hết các cấp bao gồm thêm hỗ trợ cho các hệ thống quản lý phiên bản (như Subversion) và bổ sung thêm bộ công cụ mới. như biên tập và thiết kế trực quan cho các miền ngôn ngữ cụ thể hoặc bộ công cụ dành cho các khía cạnh khác trong quy trình phát triển phần mềm.

Visual Studio hỗ trợ nhiều ngôn ngữ lập trình khác nhau và cho phép trình biên tập mã và gỡ lỗi để hỗ trợ (mức độ khác nhau) hầu như mọi ngôn ngữ lập trình. Các ngôn ngữ tích hợp gồm có C,[4] C++ và C++/CLI (thông qua Visual C++), VB.NET (thông qua Visual Basic.NET), C# (thông qua Visual C#) và F# (như của Visual Studio 2010[5]). Hỗ trợ cho các ngôn ngữ khác như J++/J#, Python và Ruby thông qua dịch vụ cài đặt riêng rẽ. Nó cũng hỗ trợ XML/XSLT, HTML/XHTML, JavaScript và CSS.

# CHƯƠNG 2: CƠ SỞ LÍ THUYẾT

## 1.Tập dữ liệu đầu vào

### 1.1 Cấu trúc dữ liệu

Tập dữ liệu đầu vào là các thông tin của trò chơi .Các thuộc tính bao gồm:

kết quả: X thắng or thua

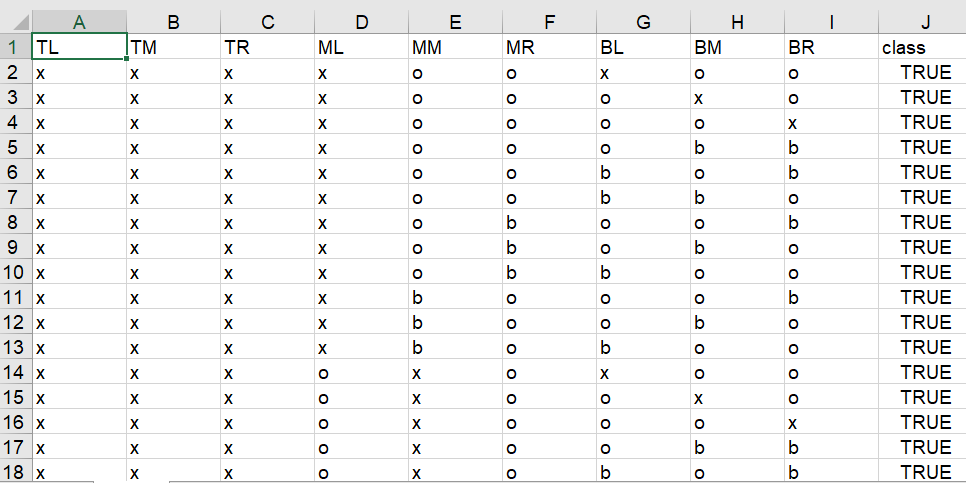
* X đánh trước
* X = 1, O = 2, trống = 0
* dữ liệu có 9 cột tương ứng 9 ô trong tic tac toe
* Kết quả: win -> X thắng, lose -> X thua.
* TL: Top Left
* TM: Top Mid
* TR: Top Right
* ML: Mid Left
* MM: Mid Mid
* MR: Mid Right
* BL: Bottom Left
* BM: bottom Mid
* BR: Bottom Right

### 1.2 Chuẩn hóa dữ liệu

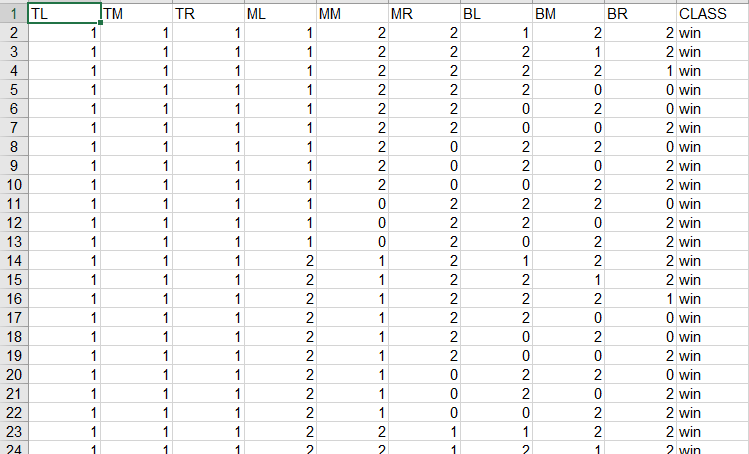
Đối với các thuộc tính có giá trị liên tục thì được biểu diễn bởi hai trạng thái là 1 hoặc 2. Đối với các ô bỏ trống thì trạng thái là 0

Kết quả quyết định được biểu diễn bằng hai giá trị là Win (nếu X thắng) và Lose (nếu X thua).

### 1.3 Dữ liệu mẫu

****

Hình 2:Dữ liệu ban đầu chưa được chuẩn hóa

**

Hình 3:Dữ liệu sau khi được chuẩn hóa

## Cây quyết định

### 2.1 Giới thiệu

Trong [lý thuyết quyết định](https://vi.wikipedia.org/wiki/L%C3%BD_thuy%E1%BA%BFt_quy%E1%BA%BFt_%C4%91%E1%BB%8Bnh), một cây quyết định ([tiếng Anh](https://vi.wikipedia.org/wiki/Ti%E1%BA%BFng_Anh): decision tree) là một [đồ thị](https://vi.wikipedia.org/wiki/L%C3%BD_thuy%E1%BA%BFt_%C4%91%E1%BB%93_th%E1%BB%8B) của các quyết định và các hậu quả có thể của nó (bao gồm rủi ro và hao phí tài nguyên). Cây quyết định được sử dụng để xây dựng một [kế hoạch](https://vi.wikipedia.org/wiki/K%E1%BA%BF_ho%E1%BA%A1ch) nhằm đạt được [mục tiêu](https://vi.wikipedia.org/wiki/M%E1%BB%A5c_ti%C3%AAu) mong muốn. Các cây quyết định được dùng để hỗ trợ quá trình ra quyết định. Cây quyết định là một dạng đặc biệt của [cấu trúc cây](https://vi.wikipedia.org/w/index.php?title=C%E1%BA%A5u_tr%C3%BAc_c%C3%A2y&action=edit&redlink=1).Cây quyết định là cấu trúc biểu diễn dưới dạng cây. Trong đó, mỗi nút trong (internal node) biễu diễn một thuộc tính, nhánh (branch) biễu diễn giá trị có thể có của thuộc tính, mỗi lá (leaf node) biểu diễn các lớp quyết định và đỉnh trên cùng của cây gọi là gốc (root). Cây quyết định có thể được dùng để phân lớp bằng cách xuất phát từ gốc của cây và di chuyển theo các nhánh cho đến khi gặp nút lá. Trên cơ sở phân lớp này chúng ta có thể chuyển đổi về các luật quyết định.

Trong lĩnh vực học máy, cây quyết định là một kiểu mô hình dự báo (predictive model), nghĩa là một ánh xạ từ các quan sát về một sự vật/hiện tượng tới các kết luận về giá trị mục tiêu của sự vật/hiện tượng. Mỗi một nút trong (internal node) tương ứng với một biến; đường nối giữa nó với nút con của nó thể hiện một giá trị cụ thể cho biến đó. Mỗi nút lá đại diện cho giá trị dự đoán của biến mục tiêu, cho trước các giá trị của các biến được biểu diễn bởi đường đi từ nút gốc tới nút lá đó. Kỹ thuật học máy dùng trong cây quyết định được gọi là học bằng cây quyết định, hay chỉ gọi với cái tên ngắn gọn là cây quyết định.

[Học bằng cây quyết định cũng là một phương pháp thông dụng trong khai phá dữ liệu. Khi đó, cây quyết định mô tả một cấu trúc cây, trong đó, các lá đại diện cho các phân loại còn cành đại diện cho các kết hợp của các thuộc tính dẫn tới phân loại đó[1]. Một cây quyết định có thể được học bằng cách chia tập hợp nguồn thành các tập con dựa theo một kiểm tra giá trị thuộc tính. Quá trình này được lặp lại một cách đệ quy cho mỗi tập con dẫn xuất. Quá trình đệ quy hoàn thành khi không thể tiếp tục thực hiện việc chia tách được nữa, hay khi một phân loại đơn có thể áp dụng cho từng phần tử của tập con dẫn xuất. Một bộ phân loại rừng ngẫu nhiên (random forest) sử dụng một số cây quyết định để có thể cải thiện tỉ lệ phân loại.

Cây quyết định cũng là một phương tiện có tính mô tả dành cho việc tính toán các xác suất có điều kiện.

Cây quyết định có thể được mô tả như là sự kết hợp của các kỹ thuật toán học và tính toán nhằm hỗ trợ việc mô tả, phân loại và tổng quát hóa một tập dữ liệu cho trước.

### 2.2 Các kiểu cây quyết định

Cây quyết định còn có hai tên khác:

* Cây hồi quy (Regression tree) ước lượng các hàm giá có giá trị là số thực thay vì được sử dụng cho các nhiệm vụ phân loại. (ví dụ: ước tính giá một ngôi nhà hoặc khoảng thời gian một bệnh nhân nằm viện)

- Cây phân loại (Classification tree), nếu y là một biến phân loại như: giới tính (nam hay nữ), kết quả của một trận đấu (thắng hay thua).

### Ưu điểm của cây quyết định

So với các phương pháp khai phá dữ liệu khác, cây quyết định là phương pháp có một số ưu điểm:

* Cây quyết định dễ hiểu. Người ta có thể hiểu mô hình cây quyết định sau khi được giải thích ngắn.
* Việc chuẩn bị dữ liệu cho một cây quyết định là cơ bản hoặc không cần thiết. Các kỹ thuật khác thường đòi hỏi chuẩn hóa dữ liệu, cần tạo các biến phụ (dummy variable) và loại bỏ các giá trị rỗng.
* Cây quyết định có thể xử lý cả dữ liệu có giá trị bằng số và dữ liệu có giá trị là tên thể loại. Các kỹ thuật khác thường chuyên để phân tích các bộ dữ liệu chỉ gồm một loại biến. Chẳng hạn, các luật quan hệ chỉ có thể dùng cho các biến tên, trong khi mạng nơ - ron chỉ có thể dùng cho các biến có giá trị bằng số.
* Cây quyết định là một mô hình hộp trắng. Mạng nơ - ron là một ví dụ về mô hình hộp đen , do lời giải thích cho kết quả quá phức tạp để có thể hiểu được.
* Có thể thẩm định một mô hình bằng các kiểm tra thống kê . Điều này làm cho ta có thể tin tưởng vào mô hình.
* Cây quyết định có thể xử lý tốt một lượng dữ liệu lớn trong thời gian ngắn. Có thể dùng máy tính cá nhân để phân tích các lượng dữ liệu lớn trong một thời gian đủ ngắn để cho phép các nhà chiến lược đưa ra quyết định dựa trên phân tích của cây quyết định.

### 3.3 Nhược điểm của cây quyết định

Kèm với đó, cây quyết định cũng có những nhược điểm cụ thể:

* Mô hình cây quyết định phụ thuộc rất lớn vào dữ liệu của bạn. Thạm chí, với một sự thay đổi nhỏ trong bộ dữ liệu, cấu trúc mô hình cây quyết định có thể thay đổi hoàn toàn.
* Cây quyết định hay gặp vấn đề overfitting.

## Thuật toán ID3

### 3.1 Giới thiệu ID3

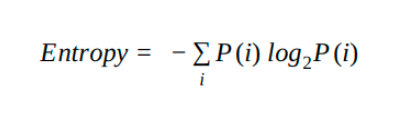
Trong [học cây quyết định](https://en.wikipedia.org/wiki/Decision_tree_learning) , ID3 ( Iterative Dichotomiser 3 ) là một [thuật toán được](https://en.wikipedia.org/wiki/Algorithm) phát minh bởi [Ross Quinlan](https://en.wikipedia.org/wiki/Ross_Quinlan)được sử dụng để tạo ra một [cây quyết định](https://en.wikipedia.org/wiki/Decision_tree_learning) từ một tập dữ liệu. ID3 là tiền thân của [thuật toán C4.5](https://en.wikipedia.org/wiki/C4.5_algorithm) và thường được sử dụng trong lĩnh vực [học máy](https://en.wikipedia.org/wiki/Machine_learning) và [xử lý ngôn ngữ tự nhiên](https://en.wikipedia.org/wiki/Natural_language_processing) .

Thuật toán ID3 bắt đầu với tập hợp ban đầu làm nút gốc . Trên mỗi lần lặp lại của thuật toán, nó sẽ lặp qua mọi thuộc tính không sử dụng của tập hợp và tính toán entropy hoặc thu được thông tin của thuộc tính đó. Sau đó, nó chọn thuộc tính có giá trị entropy nhỏ nhất (hoặc mức tăng thông tin lớn nhất). Bộ sau đó được phân tách hoặc phân vùng theo thuộc tính đã chọn để tạo ra các tập con dữ liệu. Thuật toán tiếp tục lặp lại trên mỗi tập con, chỉ xem xét các thuộc tính không bao giờ đã chọn trước đó.

Trong suốt thuật toán, cây quyết định được xây dựng với mỗi nút không phải đầu cuối (nút bên trong ) đại diện cho thuộc tính đã chọn mà dữ liệu được phân tách và các nút đầu cuối (nút lá) đại diện cho nhãn lớp của tập con cuối cùng của nhánh này.

### 3.2 Entropy

* Trong nhiệt động lực học, entropy nhiệt động lực (hay gọi đơn giản là entropy) ký hiệu là {\ displaystyle S}NS, là một đơn vị đo nhiệt năng phát tán, hấp thụ khi một hệ vật lý chuyển trạng thái tại một nhiệt độ tuyệt đối xác định {\ displaystyle T}NS ({\ displaystyle dS = dQ / T}{\ displaystyle dS = dQ / T}). Trong cơ học thống kê, entropy được định nghĩa như là một đơn vị đo lường khả năng mà một hệ có thể rơi vào trạng thái độ trong một tình trạng, nó thường được gọi là "sự lộn xộn" hay "tính bừa" thể hiện trong một
* Entropy là lượng bit trung bình tối thiểu để mã hóa thông tin khi ta biết phân bố các loại thông tin trong đó.entropy biểu thị cho sự hỗn độn, độ bất định, độ phức tạp của thông tin. Thông tin càng phức tạp càng entropy càng cao (hay công sức mã hóa lớn). Entropy nhạy cảm với thay đổi xác suất nhỏ, khi 2 phân bố càng giống nhau thì entropy càng giống nhau và ngược lạiEntropy thấp đồng nghĩa với việc hầu hết các lần nhận thông tin, ta có thể dự đoán dễ hơn, ít bất ngờ hơn, ít bất ổn hơn và ít thông tin hơn.
* Ta có công thức để tính entropy như sau:



* Trong đó P(i) là xác xuất của giá trị thứ i. Công thức này đơn giản là sự kết hợp giữa tính kích cỡ mã hóa trung bình và kích cỡ mã hóa nhỏ nhất của từng loại
* Nếu X là tập hợp tất cả các thông điệp {X1,…,Xn} mà X có thể nhận giá trị, và p(x) là xác suất X nhận giá trị x thuộc X, thì entropy của X được định nghĩa như sau

Entropy của một phép thử Bernoulli dưới dạng hàm số của xác suất thành công, thường gọi là hàm entropy nhị phân, 
  
    
      
        
          H
          
            
              b
            
          
        
        (
        p
        )
      
    
    {\displaystyle H_{\mbox{b))(p)}
  
. Entropy mỗi lần thử tối đa là 1 bit khi hai kết quả có cùng khả năng xảy ra, như trong một lần tung đồng xu công bằng.

Hình 4 : Entropy



### 3.3 Information Gain

* Information Gain dựa trên sự giảm của hàm Entropy khi tập dữ liệu được phân chia trên một thuộc tính và xác định vị trí của các nút trên cây. Để xây dựng một cây quyết định, ta phải tìm tất cả thuộc tính trả các nút về thuộc Infomation gain cao nhất.
* Để xác định được các nút nằm trong mô hình cây quyết định,chúng ta thực hiện tính Infomation Gain tại mỗi nút theo các bước sau:

+ Bước 1: Tính toán hệ số Entropy của biến mục tiêu S có N phần tử với Nc phần tử thuộc lớp c cho trước:

H(S)= – ∑cc=1 (Nc/N) log(Nc/N)

+ Bước 2: Tính hàm số Entropy tại mỗi thuộc tính: với thuộc tính x, các điểm dữ liệu trong S được chia ra K child node S1, S2, …, SK với số điểm trong mỗi child node lần lượt là m1, m2 ,…, mK , ta có:

H(x, S) = ∑Kk=1 (mk / N) \* H(Sk )

+ Bước 3: Chỉ số Gain Information được tính bằng:

G(x, S) = H(S) – H(x,S)

# CHƯƠNG 3: QUY TRÌNH XÂY DỰNG VÀ KẾT QUẢ

## Xây dựng phần mềm

### khai báo thư viện

* import pandas as pd
* from sklearn.tree import DecisionTreeClassifier
* from sklearn.model\_selection import train\_test\_split
* from sklearn import metrics
* from sklearn import tree
* from matplotlib import pyplot as plt

### đọc dữ liệu từ csv

data = pd.read\_csv("TTT.csv")

print(data.head())

### chia thuộc tính và kết quả

feature\_names = data.columns[:9]

target\_names = data['CLASS'].unique().tolist()

X = data.drop('CLASS', axis=1)

y = data['CLASS']

### chia data thành training và testing

X\_train, X\_test, y\_train, y\_test = train\_test\_split( X, y, test\_size=0.3, random\_state=1) # 70% training and 30% test

### tạo đối tượng cây quyết định loại entropy

clf = DecisionTreeClassifier(criterion="entropy")

### Huấn luyện

clf = clf.fit(X\_train, y\_train)

### dự đoán

y\_pred = clf.predict(X\_test)

### tính độ chính xác

print("Độ chính xác:", metrics.accuracy\_score(y\_test, y\_pred)\*100, "%")

### vẽ cây

fig = plt.figure(figsize=(50, 50))

\_ = tree.plot\_tree(clf, feature\_names=feature\_names, class\_names=target\_names, filled=True)

fig.savefig("decistion\_tree.png")

cây quyết định cho trò chơi tic tac toe

kết quả: X thắng or thua

X đánh trước

X = 1, O = 2, trống = 0

dữ liệu có 9 cột tương ứng 9 ô trong tic tac toe

Kết quả: win -> X thắng, lose -> X thua.

TL: Top Left

TM: Top Mid

TR: Top Right

ML: Mid Left

MM: Mid Mid

MR: Mid Right

BL: Bottom Left

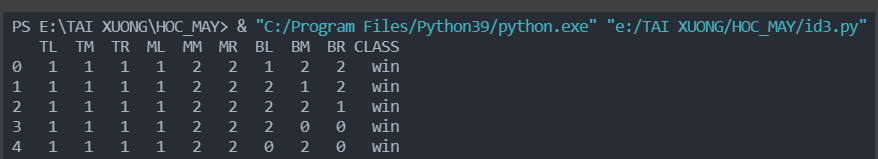
BM: bottom Mid

BR: Bottom Right

"""

## Kết quả thu được

### 2.1 Giao diện nhập dữ liệu đầu vào



Hình 5:Giao diện nhập dữ liệu đầu vào

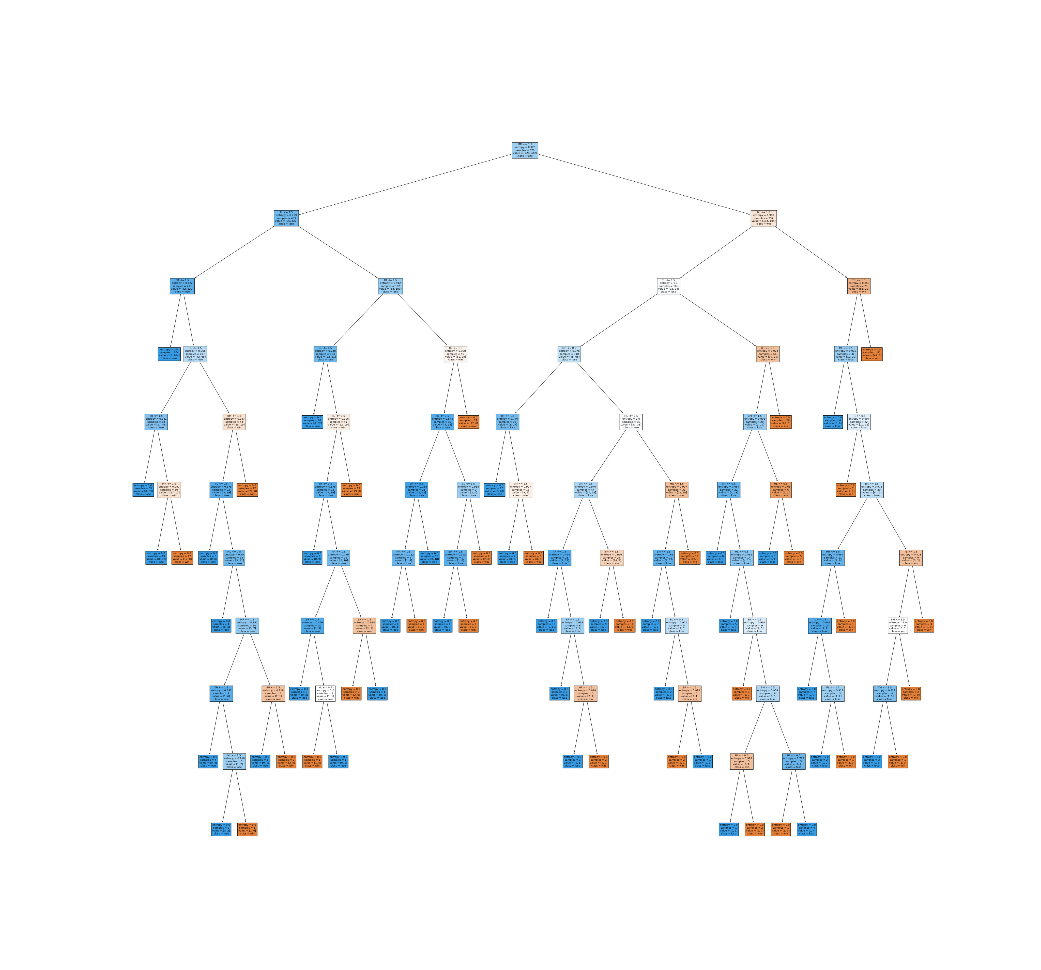
Là màn hình command line nhập hiển thị dữ liệu ra màn hình người dùng

### Giao diện kết quả

****

Hình 6: Giao diện kết quả

Giao diện hiển thị kết quả tính toán của thuật toán ID3, vẽ cây quyết định sau khi tính toán kết quả.



### 2.3 Cây quyết định của tập dữ liệu mẫu

Hình 7: Cây quyết định của tập dữ liệu mẫu

# CHƯƠNG 4: KẾT LUẬN

## Kết quả đạt được

* Hoàng thành quá trình nghiên cứu của nhóm
* Thục hiện đúng thời gian nhiệm vụ .
* Phần mềm đã thực hiện các chức năng đặt ra
* Tính được độc chính xác của cây quyết định
* Đã vẽ được cây quyết định

## Kết quả chưa đạt được và khó khăn

* Trong quá trình làm việc các dữ liệu còn khó khăn để chuyển đổi và quy ước
* Giao diện chưa thân thiện
* Do dịch bệnh nên việc họp và làm việc nhóm còn gặp nhiều khó khăn .

## Hướng phát triển

* Xây dựng giao diện theo hướng thân thiện với người dùng.
* Phát triển thêm nguồn dữ liệu đưa ra các giải quyết tối ưu cho thuật toán

# TÀI LIỆU THAM KHẢO

Tiếng Anh

1. An Introduction To Machine Learning

Tiếng Việt

1. Máy Học - Đh Bách Khoa Hn - Nguyễn Nhật Quang Website
2. <https://vi.wikipedia.org/wiki/>
3. https://medium.com/analytics-vidhya/understanding-entropy-the-golden-measurement-of-machine-learning-4ea97c663dc3