Các thành viên

|  |  |  |
| --- | --- | --- |
| **MSV** | **Họ tên** | **% tham gia** |
| B21DCCN036 | Trần Việt Dũng |  |
| B21DCCN216 | Nguyễn Trần Đạt |  |
| B21DCCN179 | Lê Văn Chiến |  |
| B21DCCN564 | Đặng Thị Hồng Ngát |  |
| B21DCCN492 | Trần Đức Lộc |  |
| B21DCCN486 | Nguyễn Thế Linh |  |

LỜI CẢM ƠN

Lời đầu tiên, em xin được gửi lời cảm ơn chân thành nhất đến thầy PTIT. Trong quá trình học tập và tìm hiểu môn Phát triển các hệ thống thông minh, em đã nhận được rất nhiều sự quan tâm, giúp đỡ, hướng dẫn tâm huyết và tận tình của thầy.

Cám ơn bạn bè

Cám ơn người thân

# MỤC LỤC

[MỤC LỤC iii](#_Toc183889054)

[DANH SÁCH CÁC HÌNH VẼ, BẢNG BIỂU v](#_Toc183889055)

[DANH SÁCH CÁC TỪ VIẾT TẮT vi](#_Toc183889056)

[Chương 1. Trí tuệ nhân tạo, hệ chuyên gia 3](#_Toc183889057)

[1.1 Trí tuệ nhân tạo 3](#_Toc183889058)

[1.1.1 Khái niệm 3](#_Toc183889059)

[1.1.2 Lịch sử và phát triển 3](#_Toc183889060)

[1.1.3 Ưu điểm và nhược điểm 5](#_Toc183889061)

[1.1.4 Ứng dụng 6](#_Toc183889062)

[1.1.5 Các lĩnh vực nghiên cứu cơ bản 8](#_Toc183889063)

[1.1.6 Các ngôn ngữ được sử dụng 10](#_Toc183889064)

[1.2 Hệ chuyên gia 11](#_Toc183889065)

[1.2.1 Định nghĩa 11](#_Toc183889066)

[1.2.2 Cấu trúc của hệ chuyên gia 12](#_Toc183889067)

[1.2.3 Ưu điểm và ứng dụng của hệ chuyên gia 13](#_Toc183889068)

[1.2.4 Xu hướng phát triển của hệ chuyên gia 14](#_Toc183889069)

[1.3 Công nghệ trí tuệ nhân tạo 15](#_Toc183889070)

[1.4 Kết luận 17](#_Toc183889071)

[Chương 2. Hệ chuyên gia dựa trên luật 18](#_Toc183889072)

[2.1 Luật, tri thức luật 18](#_Toc183889073)

[2.1.1 Luật 18](#_Toc183889074)

[2.1.2 Tri thức luật 18](#_Toc183889075)

[2.2 Tập bài toán 19](#_Toc183889076)

[2.2.1 Mô tả bài toán 19](#_Toc183889077)

[2.3 Tập luật 19](#_Toc183889078)

[2.3.1 Cây quyết định 19](#_Toc183889079)

[2.3.2 Ứng dụng cây quyết định trong hệ chuyên gia 20](#_Toc183889080)

[2.4 Kết luận 24](#_Toc183889081)

[Chương 3. Thực nghiệm 25](#_Toc183889082)

[3.1 Ngôn ngữ prolog 25](#_Toc183889083)

[3.1.1 Tổng quan về ngôn ngữ Prolog 25](#_Toc183889084)

[3.1.2 Cấu trúc của một chương trình prolog 25](#_Toc183889085)

[3.1.3 Ứng dụng của ngôn ngữ prolog 26](#_Toc183889086)

[3.2 Giao diện chương trình 27](#_Toc183889087)

[3.2.1 Giao diện chính 27](#_Toc183889088)

[Giao diện tư vấn cụ thể từng môn học 28](#_Toc183889089)

[3.4 Hỏi đáp 29](#_Toc183889090)

[3.4 Kết luận 33](#_Toc183889091)

[Kết luận 34](#_Toc183889092)

[1. Kết quả đạt được 34](#_Toc183889093)

[2. Phương hướng phát triển đề tài 34](#_Toc183889094)

[Chương 4. Phân cụm sinh viên theo điểm bằng thuật toán K-means 35](#_Toc183889095)

[4.1. Đặt bài toán 35](#_Toc183889096)

[4.2. Phân tích khám phá dữ liệu 35](#_Toc183889097)

[4.2.1 Khái niệm và mục đích sử dụng 35](#_Toc183889098)

[4.2.2 Áp dụng đối với bài toán phân cụm sinh viên 36](#_Toc183889099)

[4.2.3 Tiền xử lí dữ liệu 37](#_Toc183889100)

[4.3. Phân cụm sinh viên bằng thuật toán K-means 39](#_Toc183889101)

[4.3.1 Thuật toán K-means 39](#_Toc183889102)

[4.3.2 Xác định số cụm dựa trên phương pháp khuỷu tay 40](#_Toc183889103)

[4.3.3 Phân cụm sinh viên sử dụng thuật toán K-means 42](#_Toc183889104)

[4.4. Kết luận 44](#_Toc183889105)

[Tài liệu tham khảo 45](#_Toc183889106)

# DANH SÁCH CÁC HÌNH VẼ, BẢNG BIỂU

[Hình 1.1. Trợ lí ảo Siri của Apple 7](#_Toc183874034)

[Hình 1.2. Mô hình hệ chuyên gia 12](#_Toc183874035)

[Hình 1.3. Cấu trúc của hệ chuyên gia 13](#_Toc183874036)

[Hình 1.4. Máy tính deep blue 16](#_Toc183874037)

[Hình 2.1. Cây quyết định cho môn học “thể chất” 21](#_Toc183874038)

[Hình 2.2. Cây quyết định cho môn học “lập trình hướng đối tượng” 21](#_Toc183874039)

[Hình 2.3. Cây quyết định cho môn học “lập trình web” 22](#_Toc183874040)

[Hình 2.4. Cây quyết định cho môn học “tư tưởng Hồ Chí Minh” 22](#_Toc183874041)

[Hình 2.5. Cây quyết định cho môn học “kĩ năng thuyết trình” 23](#_Toc183874042)

[Hình 2.6. Cây quyết định cho môn học “IOT và ứng dụng” 23](#_Toc183874043)

[Hình 3.1. Giao diện chính của hệ thống 27](#_Toc183874044)

[Hình 3.2. Giao diện tư vấn cụ thể 28](#_Toc183874045)

[Hình 3.4. Nhập tên, năm học, kì học và điểm số các môn học 29](#_Toc183874046)

[Hình 3.5. Lời khuyên cho từng môn học 30](#_Toc183874047)

[Hình 3.6. Lựa chọn môn học cần tư vấn sâu 31](#_Toc183874048)

[Hình 3.7. Lựa chọn các gợi ý do hệ thống đưa ra 31](#_Toc183874049)

[Hình 3.8. Tiếp tục lựa chọn các gợi ý được đưa ra 32](#_Toc183874050)

[Hình 3.9. Hệ thống đưa ra lời khuyên cho người dùng 32](#_Toc183874051)

[Hình 3.10. Kết thúc chương trình 33](#_Toc183874052)

[Hình 4.1 Thống kê về bộ dữ liệu 36](#_Toc183874053)

[Hình 4.2 Loại bỏ các giá trị không hợp lệ 37](#_Toc183874054)

[Hình 4.3 Xử lí giá trị bị thiếu 37](#_Toc183874055)

[Hình 4.4 Biểu diễn dữ liệu dưới dạng biểu đồ 38](#_Toc183874056)

[Hình 4.5 Đồ thị hàm biến dạng của thuật toán K-Means 40](#_Toc183874057)

[Hình 4.6 Tiến hành phân cụm với k = 3 42](#_Toc183874058)

[Hình 4.7 Biểu diễn các cụm trên đồ thị 43](#_Toc183874059)

# DANH SÁCH CÁC TỪ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **Từ viết tắt** | **Tiếng Anh** | **Ý nghĩa** |
| AI | Artificial Intelligence (Trí tuệ nhân tạo) | Trí tuệ nhân tạo |
| CSDL | Database | Cơ sở dữ liệu |
| ES | Expert Systems | Hệ chuyên gia |
| IHMI | Intelligent Human-Machine Interface | Giao diện người máy thông minh |
| KB | Knowledge Base | Cơ sở tri thức |
| ABS | Anti Block Systems |  |
| IE | Inference Engine | Mô tơ suy luận |
| LISP | List Proccessing | Xử lí danh sách |
| NPL | Natural Language Processing | Xử lí ngôn ngữ tự nhiên |

MỞ ĐẦU

Trong bối cảnh phát triển nhanh chóng của cách mạng công nghiệp 4.0, Trí tuệ nhân tạo (AI) ngày càng khẳng định vai trò không thể thiếu của mình. Công nghệ này không chỉ hiện diện trong các ứng dụng phổ biến như điện thoại thông minh, trợ lý ảo mà còn được sử dụng để giải quyết các bài toán phức tạp trong các lĩnh vực như y tế, giao thông và giáo dục. AI đang dần thay đổi cách thức con người tiếp cận và giải quyết vấn đề, mang đến hiệu quả vượt trội và mở ra những cơ hội mới cho nhiều ngành nghề.

Hệ chuyên gia, một phân nhánh của AI, đã chứng minh khả năng mô phỏng trí tuệ và suy luận của con người trong các lĩnh vực cụ thể, hỗ trợ con người trong các nhiệm vụ phức tạp như chẩn đoán bệnh, quản lý tài chính và thậm chí là giáo dục. Hệ chuyên gia hoạt động dựa trên nền tảng cơ sở tri thức và các quy tắc logic để mô phỏng suy nghĩ của chuyên gia, giúp đưa ra các quyết định nhanh chóng, chính xác và có thể làm việc liên tục mà không chịu ảnh hưởng từ yếu tố cảm xúc hay mệt mỏi.

Trong quá trình học tập và nghiên cứu môn Phát triển các hệ thống thông minh, chúng em nhận thấy tiềm năng to lớn của hệ chuyên gia trong lĩnh vực giáo dục, đặc biệt là trong việc hỗ trợ sinh viên định hướng học tập. Với các hệ thống hiện nay, sinh viên phải đối mặt với nhiều thách thức trong việc lựa chọn phương pháp học tập hiệu quả, cân nhắc giữa nhiều môn học và nắm bắt các kiến thức chuyên ngành. Hệ chuyên gia mà chúng em hướng đến sẽ là một công cụ hỗ trợ đưa ra các lời khuyên học tập cá nhân hóa, giúp sinh viên có lộ trình học tập tối ưu dựa trên điểm số, kỳ học và các yếu tố cá nhân khác.

Việc xây dựng hệ chuyên gia tư vấn học tập không chỉ là cơ hội để áp dụng các lý thuyết AI vào thực tế, mà còn là một bước đi cụ thể nhằm giải quyết nhu cầu hỗ trợ cá nhân hóa trong học tập. Sự lựa chọn ngôn ngữ Prolog cũng xuất phát từ khả năng mạnh mẽ của nó trong việc xử lý các quy tắc suy luận và suy diễn dựa trên cơ sở tri thức, giúp tối ưu hóa quá trình ra quyết định của hệ thống. Hệ chuyên gia này hứa hẹn sẽ mang lại nhiều lợi ích cho sinh viên, đặc biệt là trong việc tiết kiệm thời gian tìm kiếm thông tin và nâng cao hiệu quả học tập.

Ứng dụng hệ chuyên gia vào giáo dục không chỉ giúp sinh viên mà còn hỗ trợ giáo viên và các chuyên gia giáo dục trong việc định hướng, đánh giá và cung cấp thông tin cần thiết cho quá trình học tập. Hệ thống này có thể phân tích dữ liệu học tập của sinh viên và đưa ra những gợi ý phù hợp, giúp nâng cao chất lượng giảng dạy và học tập. Các hệ chuyên gia hiện nay đã được áp dụng trong nhiều lĩnh vực như y học (chẩn đoán bệnh và tư vấn sức khỏe), tài chính (dự báo và phân tích thị trường) và hóa học (nghiên cứu cấu trúc và thành phần hóa chất). Những thành công này là minh chứng rõ ràng cho tiềm năng của hệ chuyên gia trong việc hỗ trợ ra quyết định.

Mục tiêu của nhóm là xây dựng một hệ chuyên gia tư vấn học tập cho sinh viên dựa trên ngôn ngữ Prolog, giúp đưa ra các gợi ý về phương pháp học tập, đề xuất các môn học cần chú trọng và đưa ra lời khuyên để cải thiện kết quả học tập. Để thực hiện điều này, nhóm sẽ tìm hiểu sâu về các khái niệm cơ bản của trí tuệ nhân tạo, cấu trúc hệ chuyên gia và quy trình xây dựng cây quyết định. Hệ thống tư vấn này sẽ được thử nghiệm và đánh giá dựa trên dữ liệu mẫu, với hy vọng đem lại một công cụ hữu ích, dễ sử dụng và có khả năng hỗ trợ hiệu quả cho quá trình học tập của sinh viên.

Lựa chọn đề tài này không chỉ giúp chúng em hiểu sâu hơn về trí tuệ nhân tạo và hệ chuyên gia mà còn là cơ hội để áp dụng lý thuyết vào thực tiễn, đáp ứng nhu cầu hỗ trợ học tập cá nhân hóa. Nhóm kỳ vọng hệ chuyên gia này sẽ góp phần nhỏ cải thiện chất lượng học tập, mang lại lợi ích thiết thực cho sinh viên.

# Chương 1. Trí tuệ nhân tạo, hệ chuyên gia

## 1.1 Trí tuệ nhân tạo

### 1.1.1 Khái niệm

* Trí tuệ nhân tạo là công nghệ cho phép máy móc, đặc biệt là máy tính, học hỏi và suy nghĩ như con người.
* Trí tuệ nhân tạo khác với việc lập trình logic trong các ngôn ngữ lập trình là ở việc ứng dụng các hệ thống học máy để mô phỏng trí tuệ của con người trong các xử lý mà con người làm tốt hơn máy tính.
* Trí tuệ nhân tạo giúp máy tính có được những trí tuệ của con người như: biết suy nghĩ và lập luận để giải quyết vấn đề, biết giao tiếp do hiểu ngôn ngữ, tiếng nói, biết học và tự thích nghi,…

### 1.1.2 Lịch sử và phát triển

* Khái niệm và sự quan tâm đến trí tuệ nhân tạo

Trong những năm gần đây, nhiều thuật ngữ liên quan đến trí tuệ nhân tạo như máy tính thông minh, hệ chuyên gia, mạng ngữ nghĩa,… đã xuất hiện cùng với các ngôn ngữ lập trình như Prolog và LISP.

* Lịch sử phát triển ban đầu của trí tuệ nhân tạo

Mục tiêu ban đầu: Hướng đến xây dựng các máy tính có khả năng suy nghĩ và vượt qua khả năng của con người trong một số nhiệm vụ.

Cột mốc quan trọng: Alan Turing đưa ra lý thuyết về lưu trữ chương trình trong bộ nhớ, tạo nền tảng cho máy tính hiện đại.

Các nghiên cứu tiên phong trong trí tuệ nhân tạo

* Năm 1956: Công bố chương trình dẫn xuất kết luận.
* Năm 1959: Xuất hiện các chương trình chứng minh định lý và giải quyết vấn đề tổng quát.
* Những năm 1960: Phát triển mạnh mẽ sau khi ngôn ngữ LISP được giới thiệu; Marvin Minsky đưa ra thuật ngữ "trí tuệ nhân tạo".

Thách thức trong những năm 60 và 70

* Giới hạn về phần cứng và thời gian thực thi đã cản trở việc mở rộng thành công các hệ thống trí tuệ nhân tạo.
* Vấn đề "bùng nổ tổ hợp" cản trở các tiếp cận thành công.

Thành tựu và ứng dụng thực tiễn vào cuối thập niên 70

* Tiến bộ trong xử lý ngôn ngữ tự nhiên, biểu diễn tri thức, và lý thuyết giải quyết vấn đề.
* Hệ chuyên gia bắt đầu xuất hiện trong thương mại, nổi bật với hệ MYCIN của Stanford cho y học.

Ngôn ngữ Prolog và sự mở rộng của trí tuệ nhân tạo vào thập niên 80

* Năm 1972: Prolog được giới thiệu, sau đó Nhật Bản sử dụng ngôn ngữ này cho dự án máy tính thế hệ V, thúc đẩy nghiên cứu trí tuệ nhân tạo ở Mỹ và Châu Âu.
* Từ năm 1981: Chuyển giao kết quả nghiên cứu trí tuệ nhân tạo thành ứng dụng thực tiễn và cạnh tranh thương mại hóa.
* Giai đoạn cuối thập niên 80 và đầu thập niên 90
* Sản phẩm dân dụng ứng dụng trí tuệ nhân tạo như máy giặt, máy ảnh.
* Phát triển hệ thống nhận dạng hình ảnh, tiếng nói và kỹ thuật mạng neuron.
* Quan tâm đến tiếp cận mới trong lập luận xấp xỉ và mạng neuron.

Do vậy, dù có sự chuyển biến mạnh mẽ, trí tuệ nhân tạo vẫn chưa đạt được kỳ vọng về bước nhảy vọt trong công nghệ như nhiều người mong đợi.

### 1.1.3 Ưu điểm và nhược điểm

* Ưu điểm
* Năng suất vượt trội: Tự động hóa các tác vụ lặp đi lặp lại, giải phóng sức lao động con người. Xử lý thông tin với tốc độ và khối lượng khổng lồ, vượt xa khả năng con người. Hoạt động liên tục 24/7, không cần nghỉ ngơi.
* Độ chính xác cao: Hạn chế tối đa lỗi do con người gây ra. Đưa ra quyết định dựa trên dữ liệu và thuật toán, ít bị ảnh hưởng bởi cảm xúc. Phân tích dữ liệu phức tạp để đưa ra kết quả chính xác hơn
* Khả năng sáng tạo: Hỗ trợ con người trong các lĩnh vực sáng tạo như nghệ thuật, âm nhạc, thiết kế. Tìm ra những giải pháp mới, đột phá cho các vấn đề phức tạp.
* Cá nhân hóa trải nghiệm: Cung cấp dịch vụ, sản phẩm được cá nhân hóa dựa trên sở thích và hành vi của người dùng. Nâng cao trải nghiệm của khách hàng trong nhiều lĩnh vực như mua sắm, giải trí, giáo dục.
* Giải quyết bài toán toàn cầu: Ứng dụng trong các lĩnh vực như y tế, môi trường, năng lượng để giải quyết những thách thức toàn cầu. Phân tích dữ liệu lớn để dự đoán và phòng ngừa rủi ro.
* Nhược điểm
* Nguy cơ mất việc làm: Tự động hóa có thể thay thế con người trong nhiều lĩnh vực, dẫn đến thất nghiệp. Đòi hỏi người lao động phải thích ứng và nâng cao kỹ năng để đáp ứng nhu cầu mới.
* Phân biệt đối xử: Trí tuệ nhân tạo có thể bị sai lệch do dữ liệu đầu vào thiếu khách quan hoặc mang định kiến. Gây ra những quyết định không công bằng, ảnh hưởng đến các nhóm người yếu thế.
* Vấn đề về đạo đức: Lo ngại về việc trí tuệ nhân tạo vượt khỏi tầm kiểm soát của con người. Sử dụng trí tuệ nhân tạo cho mục đích xấu như phát triển vũ khí tự động, thao túng thông tin.

### 1.1.4 Ứng dụng

Các ứng dụng của trí tuệ nhân tạo là rất đa dạng: các hệ điều khiển tự động các quá trình sản xuất công nghiệp, các robot làm việc trong các môi trường đặc biệt, các hệ chuyên gia trong các lĩnh vực, các hệ dịch tự động, các hệ nhận dạng …

* Kỹ thuật Robot: cuối những năm 1960, kỹ thuật và công nghệ chế tạo người máy có những bước phát triển mới, trên cơ sở kết hợp những thành tựu nghiên cứu của những lý thuyết trí tuệ nhân tạo và điều khiển. Các nghiên cứu và ứng dụng các phương pháp nhận dạng, hình ảnh, tiếng nói là những tiền đề quan trọng tiến tới chế tạo những robot thông minh. Người máy thế hệ thứ ba có thể đảm nhận nhiều nhiệm vụ khá phức tạp, đòi hỏi khả năng trí tuệ rất cao, các kỹ thuật trí tuệ nhân tạo cho phép điều khiển chuyển động người máy dựa trên các tri thức phụ thuộc không gian và thời gian.
* Các chương trình trò chơi: các chương trình trò chơi theo một nghĩa nào đó là xuất phát điểm của các nghiên cứu và thử nghiệm lập trình may rủi, trong đó điểm mấu chốt là xác định các hàm giá. Có rất nhiều loại chương trình trò chơi: chương trình chơi cờ, chơi bài, trò chơi Go, Nimo …
* Xử lý ngôn ngữ tự nhiên
* Giao tiếp bằng ngôn ngữ tự nhiên với các máy tự động.
* Các hệ thống thu thập tin tự động
* Robot có khả năng nghe, hiểu
* Giao tiếp với các hệ chuyên gia
* Hiểu văn bản
* Hiểu tiếng nói liên tục …
* Các hệ thống xử lý tri thức và dữ liệu tích hợp
* Xây dựng các hệ CSDL cỡ lớn và các hệ chuyên gia dựa trên tri thức dựa trên một số cách tiếp cận hợp lý cho phép xử lý cùng một lúc cả dữ liệu và tri thức, như: cách tiếp cận biểu diễn hướng đối tượng, CSDL suy diễn, Biểu diễn luật-đối tượng …
* Các hệ hỗ trợ quyết định dựa trên tri thức được xem như kết quả của sự kết hợp tri thức - dữ liệu cùng với việc sử dụng các mô hình toán học.
* Các giao diện người-máy thông minh
* Giao diện người-máy thông minh là công nghệ cải tiến giao diện nhằm giúp con người và máy móc tương tác dễ dàng, hiệu quả hơn. Một số ứng dụng cụ thể:
* Giao diện nhận diện giọng nói: Hệ thống nhận diện giọng nói cho phép người dùng tương tác với máy móc thông qua giọng nói tự nhiên, tiêu biểu là trợ lý ảo như Siri của Apple, Google Assistant, và Amazon Alexa. Các hệ thống này sử dụng các mô hình học máy và mạng nơ-ron nhân tạo để nhận diện và xử lý ngôn ngữ tự nhiên.
* Công nghệ nhận diện khuôn mặt: Sử dụng trong các thiết bị điện thoại thông minh (như Face ID của Apple) và các hệ thống an ninh, giao diện này cho phép người dùng mở khóa thiết bị hoặc truy cập vào hệ thống mà không cần mật khẩu.



Hình 1.1. Trợ lí ảo Siri của Apple

* Giao diện thông minh trong ô tô: Các hãng xe như Tesla, BMW và Mercedes-Benz trang bị các hệ thống tương tác thông minh trên xe, hỗ trợ điều khiển thông qua giọng nói hoặc cử chỉ. Những giao diện này có thể tự động điều chỉnh điều hòa, ánh sáng và thậm chí điều hướng đường đi dựa trên yêu cầu của người lái.
* Các thiết bị điện tử thông minh sử dụng logic mờ
* Điều hòa không khí thông minh: Các hãng như Mitsubishi và Daikin sử dụng logic mờ để điều chỉnh nhiệt độ dựa trên các yếu tố như nhiệt độ bên ngoài, độ ẩm và số lượng người trong phòng. Điều này giúp tối ưu hóa việc làm mát, tiết kiệm năng lượng mà vẫn đảm bảo sự thoải mái cho người dùng.
* Máy giặt thông minh: Logic mờ giúp máy giặt tự động xác định lượng nước, nhiệt độ và thời gian giặt dựa trên khối lượng, loại vải và độ bẩn của đồ giặt. Một ví dụ nổi bật là máy giặt của hãng LG với tính năng TurboWash, cho phép giặt nhanh hơn mà vẫn sạch.
* Hệ thống phanh trong ô tô: Các hệ thống phanh như ABS sử dụng logic mờ để dự đoán và điều chỉnh lực phanh trong các điều kiện trơn trượt, giúp xe dừng an toàn mà không bị trượt bánh.

### 1.1.5 Các lĩnh vực nghiên cứu cơ bản

* Lý thuyết giải quyết vấn đề và các kỹ thuật suy diễn thông minh
* Phương pháp biểu diễn bài toán trong không gian và chiến lược tìm kiếm đồ thị trạng thái
* Phương pháp này bắt đầu bằng việc xác định không gian trạng thái, nơi mỗi trạng thái đại diện cho một tình huống cụ thể trong bài toán. Bằng cách mô hình hóa bài toán như một đồ thị, các đỉnh đại diện cho các trạng thái và các cạnh thể hiện các chuyển đổi giữa các trạng thái. Chiến lược tìm kiếm sau đó được áp dụng để điều hướng qua đồ thị trạng thái nhằm tìm ra giải pháp tối ưu hoặc một giải pháp chấp nhận được.
* Các thuật toán tìm kiếm như tìm kiếm theo chiều sâu, tìm kiếm theo chiều rộng, và thuật toán A\* thường được sử dụng trong phương pháp này.
* Phương pháp quy bài toán về các bài toán con và chiến lược tìm kiếm trên đồ thị và/hoặc:
* Phương pháp quy bài toán về các bài toán con dựa trên nguyên lý chia để trị. Nó chia nhỏ bài toán lớn thành các bài toán con nhỏ hơn, dễ giải quyết hơn. Mỗi bài toán con này được giải quyết độc lập và sau đó kết hợp các giải pháp lại với nhau để có được giải pháp cho bài toán gốc.
* Các chiến lược tìm kiếm trên đồ thị có thể được áp dụng để tìm ra các giải pháp cho từng bài toán con, giúp tối ưu hóa quá trình giải quyết và giảm thiểu độ phức tạp của bài toán.
* Phương pháp GPS: GPS là một phương pháp nổi bật trong trí tuệ nhân tạo, được phát triển để giải quyết các bài toán tổng quát thông qua việc sử dụng các quy tắc suy diễn. Nó hoạt động bằng cách tìm kiếm một chuỗi các trạng thái từ trạng thái hiện tại đến trạng thái mục tiêu, sử dụng các quy tắc để điều hướng giữa các trạng thái.
* Phương pháp hình thức sử dụng cách tiếp cận logic: sử dụng logic để biểu diễn và giải quyết vấn đề. Các bài toán được mô hình hóa dưới dạng các biểu thức logic, cho phép áp dụng các quy tắc logic để suy diễn và tìm ra giải pháp. Cách tiếp cận logic không chỉ cung cấp một khung lý thuyết vững chắc để phân tích vấn đề mà còn cho phép tự động hóa quy trình suy diễn, giúp máy tính có thể đưa ra quyết định dựa trên các quy tắc đã được xác định.
* Một số hệ thống dựa trên logic như hệ thống lập luận, Prolog, và các hệ thống logic mờ cũng thường được sử dụng để phát triển các ứng dụng trí tuệ nhân tạo.
* Lý thuyết tìm kiếm may rủi.
* Lập trình may rủi là một hướng tiếp cận quan trọng trong việc xây dựng các hệ thống trí tuệ nhân tạo.
* Lý thuyết tìm kiếm may rủi bao gồm các phương pháp và các kỹ thuật tìm kiếm, sử dụng các tri thức đặc biệt hay sinh từ bản thân bài toán cần giải để rút ngắn quá trình giải, nhanh chóng đi đến kết quả mong muốn.
* Kỹ thuật cơ bản dựa trên các giao thức may rủi hay được sử dụng trong thực tiễn là các hàm đánh giá.
* Lý thuyết biểu diễn tri thức và kỹ nghệ xử lý tri thức
* Hai thành phần cơ bản trong mọi hệ thống trí tuệ nhân tạo là các phương pháp biểu diễn tri thức và các phương pháp suy diễn, các phương pháp tìm kiếm tương ứng.
* Các kỹ nghệ xử lý tri thức đang là lĩnh vực đang được tập trung nghiên cứu của các chuyên gia trí tuệ nhân tạo. Sự ra đời của các hệ chuyên gia là kết quả của sự kết hợp những tư tưởng cơ bản trong biểu diễn tri thức và các phương pháp suy diễn. Những năm gần đây, người ta thường quan tâm đến việc xử lý tri thức và dữ liệu tích hợp, tạo tiền đề cho những áp dụng thực tế của các hệ trợ giúp quyết định.
* Lý thuyết nhận dạng
* Vấn đề nhận dạng thông tin hình ảnh, tiếng nói vẫn còn là một thách thức đối với các chuyên gia trí tuệ nhân tạo. Nhờ công cụ hệ chuyên gia, người ta đã xây dựng được một số hệ thống trí tuệ nhân tạo được hình ảnh ba chiều, hiểu được tiếng nói liên tục dựa trên tri thức về ảnh, ngữ cảnh và ngữ điệu giọng nói.
* Cơ sở toán học của lý thuyết nhận dạng được xây dựng và phát triển theo các cách tiếp cận: lý thuyết thống kê về nhận dạng, lý thuyết cấu trúc về nhận dạng, lý thuyết đại số về nhận dạng, lý thuyết may rủi về nhận dạng.

### 1.1.6 Các ngôn ngữ được sử dụng

Python

* Ưu điểm: Python được sử dụng phổ biến trong AI nhờ cú pháp đơn giản và dễ học, cộng với hệ sinh thái thư viện phong phú cho AI và học máy, như TensorFlow, Keras, PyTorch, Scikit-Learn, và NLTK.
* Ứng dụng: học máy, thị giác máy tính, xử lý ngôn ngữ tự nhiên, và phân tích dữ liệu.
* Nhược điểm: hiệu suất thấp hơn so với các ngôn ngữ biên dịch như C++.

Prolog

* Ưu điểm: Prolog là một ngôn ngữ dựa trên logic, rất phù hợp cho các hệ thống chuyên gia, phân tích cú pháp, và suy luận logic.
* Ứng dụng: các hệ chuyên gia, xử lý ngôn ngữ tự nhiên, và suy luận logic.
* Nhược điểm: khó tiếp cận cho người mới và ít phổ biến trong các ứng dụng AI hiện đại.

LISP

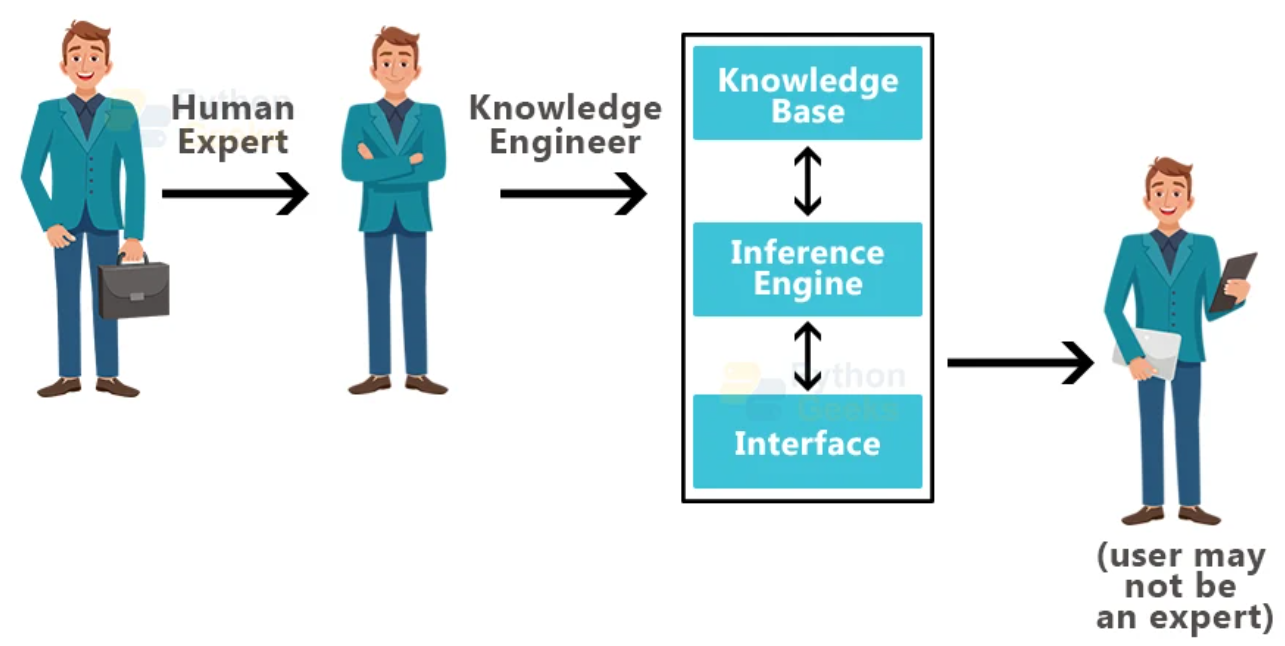
* Ưu điểm: LISP là ngôn ngữ AI đầu tiên, được thiết kế linh hoạt và hỗ trợ cho các biểu thức đệ quy. Nó thích hợp cho việc phát triển các ứng dụng AI cần linh hoạt cao.
* Ứng dụng: hệ chuyên gia, xử lý ngôn ngữ tự nhiên, và học máy.
* Nhược điểm: không được sử dụng phổ biến và cú pháp phức tạp hơn so với Python.

## 1.2 Hệ chuyên gia

### 1.2.1 Định nghĩa

Hệ chuyên gia là chương trình máy tính được thiết kế để mô hình hoá khả năng giải quyết vấn đề của người chuyên gia.

Hệ chuyên gia còn được định nghĩa như sau: hệ chuyên gia là một hệ thống dựa trên tri thức, cho phép mô hình hóa các tri thức của chuyên gia có chất lượng, hay chuyên gia được huấn luyện trong một lĩnh vực cụ thể, cho phép dùng tri thức này để giải quyết vấn đề phức tạp thuộc lĩnh vực này.

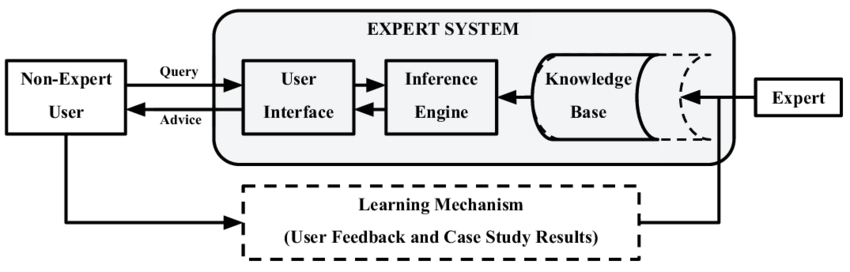


Hình 1.2. Mô hình hệ chuyên gia

### 1.2.2 Cấu trúc của hệ chuyên gia

Hệ chuyên gia bao gồm yếu tố quan trọng là tri thức chuyên gia và lập luận. Hệ thống bao gồm 2 khối chính là cơ sở tri thức và mô-tơ suy luận.

* Cơ sở tri thức: Chứa các tri thức chuyên sâu về lĩnh vực như chuyên gia. Cơ sở này gồm các sự kiện, các luật, các khái niệm và các quan hệ. Người ta cần thể hiện các tri thức này ở dạng thích hợp.
* Lưu trữ tri thức từ các chuyên gia trong một lĩnh vực nhất định.
* Bao gồm các luật ngẫu nhiên và tri thức trực cảm được tích lũy qua kinh nghiệm.
* Dữ liệu tạm thời: Cơ sở sự kiện chứa thông tin từ người dùng, cập nhật khi hệ thống hoạt động.
* Mô-tơ suy luận: Là bộ xử lý tri thức được mô hình hoá theo cách lập luận của chuyên gia. Mô-tơ hoạt động trên thông tin về vấn đề đang xét, so sánh với tri thức lưu trong cơ sở tri thức rồi rút ra kết luận hay bình luận. Như vậy, người ta cần có kỹ thuật về suy diễn.



Hình 1.3. Cấu trúc của hệ chuyên gia

### 1.2.3 Ưu điểm và ứng dụng của hệ chuyên gia

* Khả năng vượt trội:
* Tái hiện khả năng chuyên gia: Hệ chuyên gia được phát triển dựa trên kiến thức và kinh nghiệm của các chuyên gia trong từng lĩnh vực cụ thể. Nhờ vào các thuật toán suy luận và hệ thống kiến thức được tích hợp, hệ chuyên gia có thể thực hiện các phân tích và đưa ra những quyết định phức tạp một cách nhanh chóng, giúp tiết kiệm thời gian và nâng cao hiệu quả so với con người.
* Làm việc liên tục và không chịu ảnh hưởng cảm xúc: Khác với con người, hệ chuyên gia có thể hoạt động liên tục 24/7 mà không mệt mỏi hay bị ảnh hưởng bởi các yếu tố cảm xúc như căng thẳng, mệt mỏi hay thiên kiến. Điều này giúp duy trì tính nhất quán trong quá trình xử lý thông tin và ra quyết định.
* Khả năng học hỏi và cải tiến: Một số hệ chuyên gia hiện đại có khả năng học từ các dữ liệu mới để liên tục cải tiến kiến thức và nâng cao khả năng suy luận. Điều này giúp hệ chuyên gia thích nghi với các thay đổi trong môi trường ứng dụng và trở nên ngày càng chính xác và thông minh hơn.
* Ứng dụng trong thực tế
* Y tế: Hệ chuyên gia giúp hỗ trợ chẩn đoán bệnh, đề xuất phác đồ điều trị, theo dõi tình trạng bệnh nhân và dự báo tình trạng sức khỏe. Nhờ đó, hệ chuyên gia góp phần giảm tải công việc cho đội ngũ y bác sĩ và nâng cao chất lượng chăm sóc bệnh nhân.
* Hóa học: Trong ngành hóa học, hệ chuyên gia hỗ trợ phân tích cấu trúc các hợp chất, tìm kiếm các thành phần tối ưu cho dược phẩm, và tự động hóa các quy trình phức tạp trong nghiên cứu và sản xuất.
* Tài chính: Hệ chuyên gia có khả năng phân tích dữ liệu tài chính, đưa ra dự báo thị trường, phát hiện gian lận và đề xuất các chiến lược đầu tư hiệu quả. Các hệ thống này giúp nâng cao hiệu quả kinh doanh và bảo vệ lợi ích của các tổ chức tài chính.
* Khai thác khoáng sản: Hệ chuyên gia như prospector giúp xác định vị trí các mỏ khoáng sản tiềm năng dựa trên dữ liệu địa chất, tiết kiệm chi phí và công sức trong quá trình khảo sát và khai thác tài nguyên.

### 1.2.4 Xu hướng phát triển của hệ chuyên gia

Tích hợp với hệ thống khác:

* Kết hợp hệ chuyên gia với các hệ thống thông tin truyền thống: Việc tích hợp hệ thống hệ chuyên gia với các hệ thống thông tin truyền thống là một bước quan trọng để đảm bảo rằng hệ thống có thể hoạt động hiệu quả trong một môi trường đa dạng và phức tạp. Bằng cách kết nối hệ chuyên gia với các hệ thống quản lý hiện có, như hệ thống quản lý khách hàng, hệ thống kế toán, và các phần mềm quản lý khác, hệ chuyên gia có thể tận dụng được dữ liệu và quy trình từ những hệ thống này.
* Khi tích hợp, các dữ liệu như thông tin khách hàng, lịch sử giao dịch, và thông tin đặt chỗ có thể được đồng bộ hóa giữa các hệ thống, giúp cải thiện quy trình ra quyết định và cung cấp dịch vụ tốt hơn cho khách hàng. Ví dụ, thông tin về khách hàng từ hệ thống có thể được sử dụng để cá nhân hóa trải nghiệm người dùng trong hệ chuyên gia, trong khi dữ liệu từ hệ thống kế toán có thể cung cấp thông tin chi tiết về doanh thu và chi phí.
* Kết nối với cơ sở dữ liệu: Hệ thống hệ chuyên gia cần được kết nối với một cơ sở dữ liệu mạnh mẽ để lưu trữ và truy xuất dữ liệu một cách hiệu quả. Việc sử dụng các hệ quản trị cơ sở dữ liệu như MySQL, PostgreSQL, hoặc Oracle cho phép hệ chuyên gia lưu trữ thông tin về phim, suất chiếu, và thông tin khách hàng một cách có tổ chức và an toàn. Hệ thống cần đảm bảo rằng việc truy vấn dữ liệu diễn ra nhanh chóng và mượt mà, đồng thời hỗ trợ các thao tác như thêm, sửa, xóa dữ liệu một cách dễ dàng.
* Việc kết nối với cơ sở dữ liệu cũng cho phép hệ chuyên gia thực hiện các phân tích dữ liệu nâng cao. Các công cụ phân tích dữ liệu có thể được tích hợp để giúp phân tích xu hướng đặt vé, thói quen của khách hàng, và hiệu suất của từng bộ phim. Điều này không chỉ giúp quản lý ra quyết định thông minh hơn mà còn cung cấp những thông tin giá trị để tối ưu hóa kế hoạch marketing và cải thiện trải nghiệm khách hàng.
* Hệ thống phân tích dữ liệu: Để tối đa hóa giá trị từ dữ liệu thu thập được, việc tích hợp hệ chuyên gia với các hệ thống phân tích dữ liệu là rất quan trọng. Các công cụ phân tích như Tableau, Power BI, hoặc Google Data Studio có thể được sử dụng để tạo ra các báo cáo trực quan và dashboard dễ hiểu. Điều này giúp đội ngũ quản lý nắm bắt nhanh chóng tình hình kinh doanh và có những điều chỉnh cần thiết trong chiến lược hoạt động.

## 1.3 Công nghệ trí tuệ nhân tạo

Có nhiều cách phân loại trí tuệ nhân tạo thành các nhóm. Trong đó, phổ biến hơn cả là phân loại công nghệ trí tuệ nhân tạo, gồm có 4 loại như sau:

* Công nghệ trí tuệ nhân tạo phản ứng
* Công nghệ trí tuệ nhân tạo phản ứng có khả năng dự đoán và phân tích những hành động của chính mình và của người đối diện, từ đó có phản ứng hợp lý nhất.
* Một ví dụ điển hình của công nghệ trí tuệ nhân tạo phản ứng là Deep Blue. Đây là một chương trình chơi cờ vua tự động, được tạo ra bởi IBM, với khả năng xác định các nước cờ đồng thời dự đoán những bước đi tiếp theo của đối thủ. Thông qua đó, Deep Blue đưa ra những nước đi thích hợp nhất.



Hình 1.4. Máy tính deep blue

* Ngoài ra còn một sản phẩm khác do Google chế tạo – gọi là [AlphaGO](https://quantrimang.com/cuoc-song/tim-hieu-tri-tue-nhan-tao-alphago-khien-con-nguoi-than-phuc-121210) – một ứng dụng giải trí chơi cờ vây trên máy tính.
* Công nghệ trí tuệ nhân tạo với bộ nhớ hạn chế
* Công nghệ trí tuệ nhân tạo bộ nhớ hạn chế tiến bộ hơn so với công nghệ trí tuệ nhân tạo phản ứng ở chỗ nó có thể dựa vào những kinh nghiệm trong quá khứ để áp dụng đưa ra các hành động phù hợp hơn trong tương lai.
* Ứng dụng phổ biến: máy bay drone, tàu ngầm, xe hơi không người lái
* Công nghệ trí tuệ nhân tạo này được kết hợp với cảm biến của môi trường xung quanh để dự đoán những tình huống có thể xảy ra và đưa ra những quyết định tốt nhất cho thiết bị.
* Ví dụ như đối với xe không người lái, nhiều cảm biến được trang bị xung quanh xe và ở đầu xe để tính toán khoảng cách với các xe phía trước, công nghệ trí tuệ nhân tạo sẽ dự đoán khả năng xảy ra va chạm, từ đó điều chỉnh tốc độ xe phù hợp để giữ an toàn cho xe.
* Lý thuyết trí tuệ nhân tạo
* Công nghệ trí tuệ nhân tạo này có thể học hỏi cũng như tự suy nghĩ, sau đó áp dụng những gì học được để thực hiện một việc cụ thể. Hiện nay, công nghệ trí tuệ nhân tạo này vẫn chưa trở thành một phương án khả thi.
* Một ví dụ điển hình là sản phẩm của Facebook nhằm hỗ trợ quá trình giao tiếp bằng công nghệ. Dự án trí tuệ nhân tạo này vượt quá khả năng của đội ngũ trí tuệ nhân tạo thời điểm đó, song không thể phủ nhận sự phát triển vượt bậc của loại hình này.
* Công nghệ trí tuệ nhân tạo tự nhận thức
* Công nghệ trí tuệ nhân tạo này có khả năng tự nhận thức về bản thân, có ý thức và hành xử như con người. Thậm chí, chúng còn có thể bộc lộ cảm xúc cũng như hiểu được những cảm xúc của con người. Đây được xem là bước phát triển cao nhất của công nghệ trí tuệ nhân tạo và đến thời điểm hiện tại, công nghệ này vẫn chưa khả thi.

## 1.4 Kết luận

Trong chương này, nhóm đã làm rõ các khái niệm căn bản về trí tuệ nhân tạo và hệ chuyên gia, bao gồm đặc điểm, khả năng ứng dụng, và tiềm năng của AI trong việc giải quyết các vấn đề phức tạp. Đặc biệt, hệ chuyên gia được phân tích chi tiết như một công nghệ cốt lõi của AI, thể hiện vai trò quan trọng trong việc mô phỏng tri thức và suy luận của con người.

Thông qua sự kết hợp của AI và hệ chuyên gia, các hệ thống thông minh có thể học hỏi và đưa ra quyết định tương tự như con người, hỗ trợ giải quyết các vấn đề trong nhiều lĩnh vực như y tế, tài chính, giáo dục, và sản xuất. Công nghệ AI không chỉ giúp cải thiện hiệu suất làm việc mà còn mở ra nhiều cơ hội mới cho nghiên cứu và phát triển các ứng dụng tiên tiến.

Nền tảng về AI và hệ chuyên gia được thiết lập làm cơ sở cho việc nghiên cứu chi tiết các hệ chuyên gia dựa trên luật, giúp xây dựng các hệ thống có khả năng suy luận tự động và hỗ trợ người dùng ra quyết định trong các tình huống cụ thể.

# Chương 2. Hệ chuyên gia dựa trên luật

## 2.1 Luật, tri thức luật

### 2.1.1 Luật

Định nghĩa**:** Luật trong hệ chuyên gia thường được biểu diễn dưới dạng các quy tắc "Nếu - Thì" (IF-THEN). Mỗi quy tắc gồm hai phần:

* Phần điều kiện (IF): Các điều kiện cần được thỏa mãn.
* Phần kết luận (THEN): Hành động hoặc kết quả được thực hiện nếu các điều kiện đó được thỏa mãn.

Chức năng**:** Luật giúp hệ thống quyết định cách thức xử lý thông tin đầu vào để đưa ra các kết luận hoặc khuyến nghị, cho phép hệ thống mô phỏng quá trình ra quyết định của con người.

Ví dụ:

* IF điểm Toán >= 8 AND điểm Lý >= 7 THEN khuyên nên chọn khối A.
* IF điểm Văn < 5 THEN khuyên nên ôn tập môn Văn.

### 2.1.2 Tri thức luật

Tri thức luật là liên kết hai mệnh đề A, B sao cho A = 1 thì B = 1

Ví dụ: Trán rộng →Thông minh, A trán rộng ⇒ A thông minh

Cấu trúc và các thành phần chính của tri thức luật trong hệ chuyên gia bao gồm:

Luật suy diễn

Định dạng: Các luật suy diễn thường được biểu diễn dưới dạng câu "Nếu-Thì" (IF-THEN), nơi mà điều kiện (IF) được đáp ứng, thì hành động hoặc kết luận (THEN) sẽ được thực hiện.

* Chức năng: Luật suy diễn cho phép hệ chuyên gia đưa ra các kết luận hoặc hành động dựa trên các sự kiện hoặc dữ liệu đầu vào.
* Ví dụ: Trong một hệ chuyên gia y khoa, có luật: "Nếu bệnh nhân có triệu chứng sốt cao và đau nhức cơ, thì có khả năng mắc bệnh cúm."

Mô hình và cơ chế suy luận

* Cơ chế suy luận: Xác định cách thức mà hệ chuyên gia sẽ áp dụng các luật suy diễn để đưa ra kết luận.
* Suy luận tiến: Hệ chuyên gia áp dụng liên tiếp các luật suy diễn từ dữ liệu hiện có, từng bước đưa ra kết luận mới cho đến khi đạt được mục tiêu.
* Suy diễn lùi: Hệ chuyên gia bắt đầu từ mục tiêu cuối cùng và truy ngược để kiểm tra xem liệu các điều kiện có được đáp ứng không, từ đó xác định tính khả thi của kết luận.

## 2.2 Tập bài toán

### 2.2.1 Mô tả bài toán

Yêu cầu với hệ thống: người dùng nhập các thông tin cá nhân, điểm số các môn, hệ thống sẽ đưa ra lời khuyên cụ thể cho người dùng dựa trên mô hình cây quyết định.

Input

* Họ tên sinh viên.
* Năm học (Ví dụ: Năm 1, Năm 2, Năm 3, Năm 4).
* Kỳ học (Ví dụ: Kỳ 1, Kỳ 2).
* Điểm số các môn học.

Output

* Đưa ra lời khuyên tổng quát cho từng môn học.
* Lời khuyên chuyên sâu cho môn học.

## 2.3 Tập luật

### 2.3.1 Cây quyết định

Cây quyết định là một công cụ phổ biến trong các bài toán phân loại và dự đoán, giúp đưa ra quyết định dựa trên các tiêu chí cụ thể. Được biểu diễn dưới dạng cấu trúc cây, mỗi nút của cây đại diện cho một thuộc tính hoặc điều kiện mà hệ thống cần kiểm tra, các nhánh đại diện cho kết quả của điều kiện này, và các nút lá thể hiện kết quả cuối cùng hoặc hành động cần thực hiện. Cấu trúc này tương tự quá trình con người suy nghĩ và ra quyết định dựa trên các tiêu chí cụ thể.

Một cây quyết định là một cấu trúc thuật toán giống như sơ đồ khối, với mỗi một nút là một phép so sánh, nhằm thực hiện thử một đặc trưng nào đó. Mỗi một nhánh thể hiện kết quả của phép so sánh đó, và mỗi một ngọn, tức nút cuối, biểu diễn kết quả phân loại của thuật toán, sau khi tính . Đường đi từ gốc đến ngọn thể hiện luật phân loại.

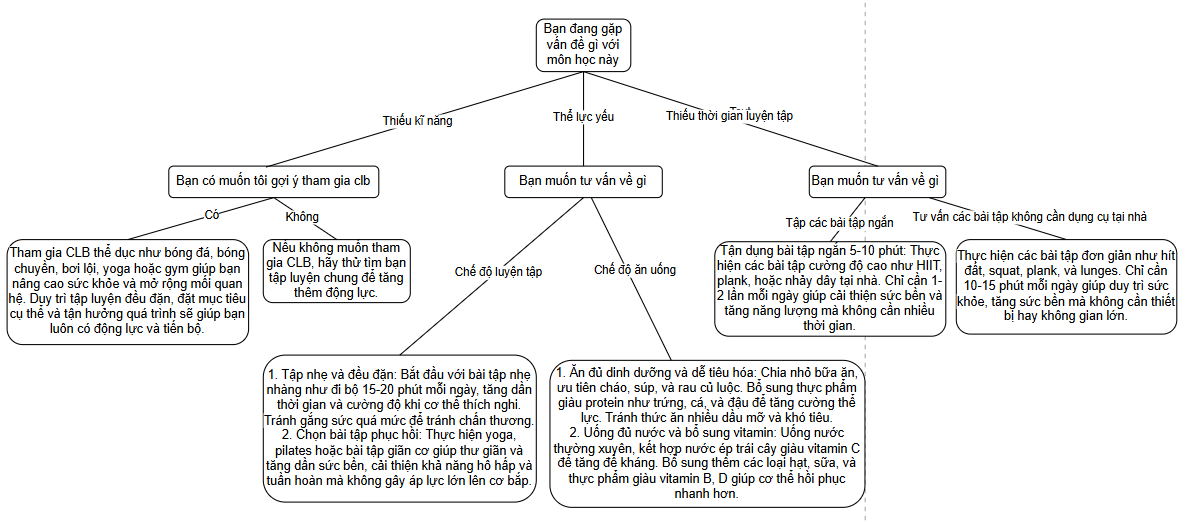
Cây quyết định có một số ưu điểm vượt trội như: dễ hiểu và giải thích, trực quan, không yêu cầu kiến thức chuyên sâu về toán học hay xác suất, và phù hợp với dữ liệu rời rạc và liên tục.

### 2.3.2 Ứng dụng cây quyết định trong hệ chuyên gia

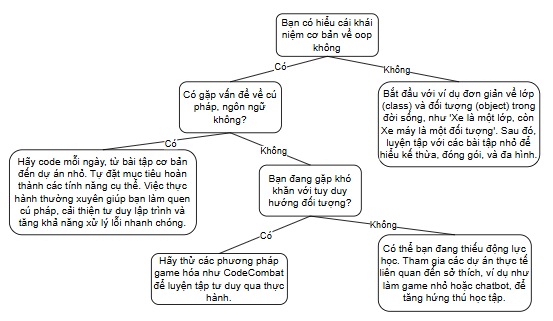
Hệ chuyên gia là một hệ thống trí tuệ nhân tạo mô phỏng khả năng suy luận và đưa ra quyết định như một chuyên gia trong lĩnh vực cụ thể. Cây quyết định được ứng dụng trong các hệ chuyên gia để thực hiện các phép suy diễn và phân tích dữ liệu phức tạp nhằm đưa ra kết quả hoặc đề xuất cụ thể. Trong hệ chuyên gia, cây quyết định giúp tạo ra các quy tắc suy luận dựa trên dữ liệu đầu vào của người dùng. Các quy tắc này được xây dựng dựa trên kiến thức của các chuyên gia trong lĩnh vực đó và tổ chức dưới dạng cây, giúp hệ thống ra quyết định một cách nhanh chóng và chính xác.

Trong trường hợp trên, các cây quyết định được xây dựng để hỗ trợ sinh viên trong quá trình học tập và rèn luyện kỹ năng. Dựa trên câu trả lời của sinh viên về điểm số hiện tại, hệ thống sẽ đi qua các nút trong cây quyết định, mỗi nút thể hiện một câu hỏi hoặc điều kiện cụ thể, nhằm thu hẹp phạm vi để tìm ra phương pháp học tập phù hợp nhất cho từng cá nhân.

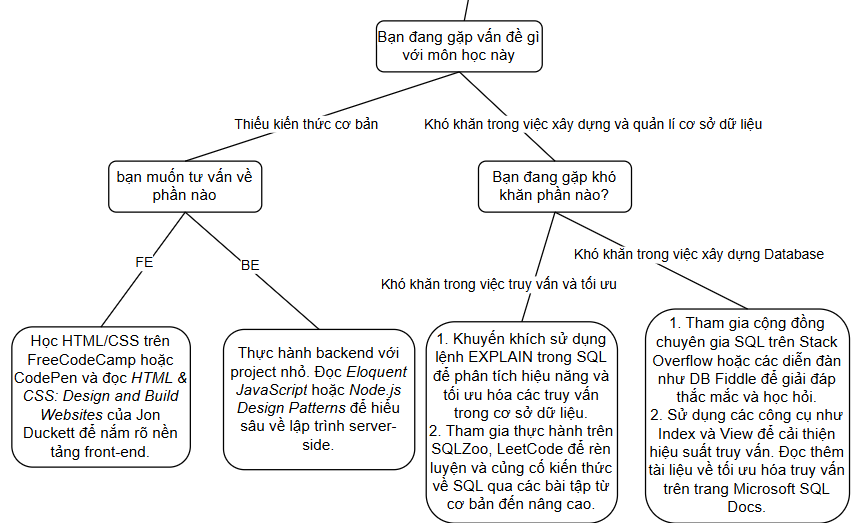
Xây dựng cây quyết định cho từng môn học cụ thể



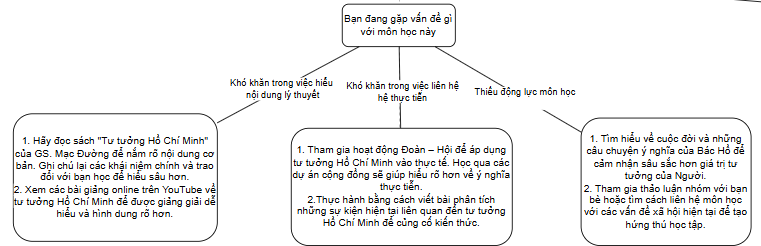
Hình 2.1. Cây quyết định cho môn học “thể chất”



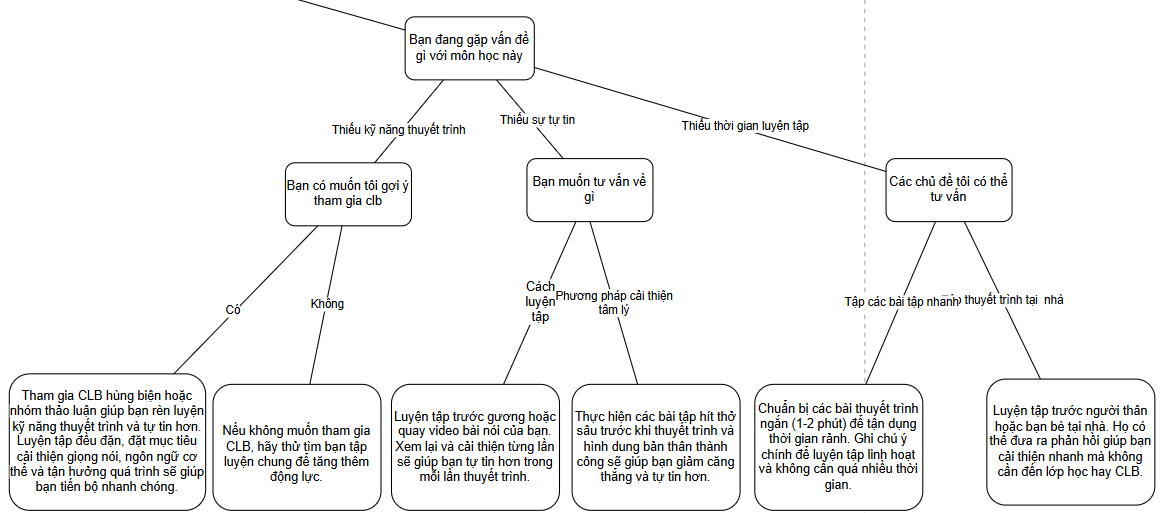
Hình 2.2. Cây quyết định cho môn học “lập trình hướng đối tượng”



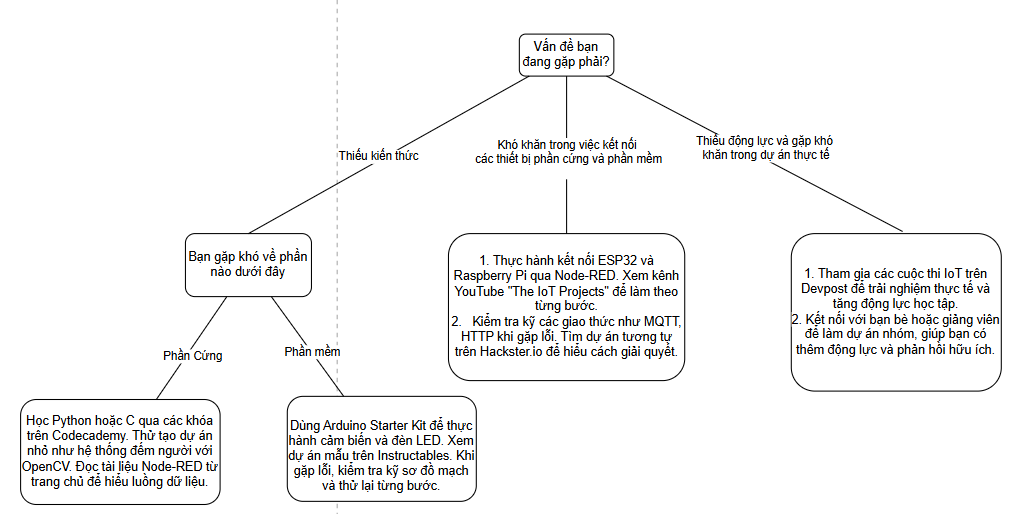
Hình 2.3. Cây quyết định cho môn học “lập trình web”



Hình 2.4. Cây quyết định cho môn học “tư tưởng Hồ Chí Minh”



Hình 2.5. Cây quyết định cho môn học “kĩ năng thuyết trình”



Hình 2.6. Cây quyết định cho môn học “IOT và ứng dụng”

## 2.4 Kết luận

Trong chương này, đã làm rõ các khái niệm cơ bản liên quan đến luật, tri thức luật và cây quyết định trong hệ chuyên gia. Luật, đặc biệt dưới dạng quy tắc "Nếu - Thì", là nền tảng của các hệ chuyên gia, giúp hệ thống ra quyết định dựa trên các điều kiện cụ thể. Tri thức luật trong hệ chuyên gia chứa đựng các quy tắc này, cho phép hệ thống mô phỏng khả năng suy luận của con người.

Cây quyết định đóng vai trò quan trọng trong hệ thống này nhờ khả năng phân loại và dự đoán, giúp tổ chức các quy tắc dưới dạng cấu trúc cây, cho phép hệ thống tiếp cận một cách trực quan và hiệu quả với các bài toán suy luận. Được ứng dụng trong nhiều lĩnh vực, cây quyết định giúp các hệ chuyên gia đưa ra các khuyến nghị và phân tích chính xác dựa trên dữ liệu đầu vào.

Ứng dụng vào hệ thống hỗ trợ học tập, cây quyết định đóng góp quan trọng bằng cách giúp sinh viên xác định lộ trình học tập phù hợp dựa trên điểm số và thông tin cá nhân. Cấu trúc cây cho phép hệ thống đi sâu vào từng điều kiện cụ thể, từ đó đề xuất các phương pháp học tập và tài liệu ôn tập phù hợp với năng lực của mỗi cá nhân.

Có thể thấy cây quyết định và luật "Nếu - Thì" không chỉ là những công cụ lý thuyết mà còn mang tính ứng dụng, giúp hệ chuyên gia nâng cao hiệu quả và tính chính xác trong việc hỗ trợ ra quyết định. Các quy tắc suy luận và cấu trúc cây này mang lại lợi ích cho người dùng, tạo ra một môi trường học tập và phát triển kỹ năng cá nhân hóa.

# Chương 3. Thực nghiệm

## 3.1 Ngôn ngữ prolog

### 3.1.1 Tổng quan về ngôn ngữ Prolog

Prolog là một ngôn ngữ lập trình mô tả. Tên gọi Prolog được xuất phát từ cụm từ tiếng Pháp *Programmation en logique*, nghĩa là "lập trình theo lô gíc". Xuất hiện từ năm 1972 do [Alain Colmerauer](https://vi.wikipedia.org/w/index.php?title=Alain_Colmerauer&action=edit&redlink=1) và [Robert Kowalski](https://vi.wikipedia.org/w/index.php?title=Robert_Kowalski&action=edit&redlink=1) phát triển), mục tiêu của Prolog là giúp người dùng mô tả lại bài toán trên ngôn ngữ của logic, dựa trên đó, máy tính sẽ tiến hành suy diễn tự động dựa vào những cơ chế suy diễn có sẵn để đưa ra câu trả lời cho người dùng.

### 3.1.2 Cấu trúc của một chương trình prolog

Một chương trình Prolog bao gồm ba thành phần chính: sự kiện, luật, và truy vấn.

Sự kiện

* Sự kiện là các mệnh đề mô tả một số thông tin hoặc quan hệ giữa các đối tượng.
* Cú pháp của một sự kiện là: predicate(argument1, argument2, ..., argumentN).
* Ví dụ: likes(lan, pho). % Lan thích Phở.

Luật

* Luật được sử dụng để định nghĩa các mối quan hệ phức tạp hơn, có điều kiện và cho phép suy diễn các tri thức mới từ những tri thức đã có.
* Cú pháp: head :- body.
* Ví dụ: cham\_chi(HocSinh) :- diem\_thap(HocSinh). -> học sinh điểm thấp thì nên học chăm chỉ.

Truy vấn

* Truy vấn được sử dụng để kiểm tra tính đúng đắn của các mệnh đề, hoặc tìm kiếm thông tin trong cơ sở tri thức. Khi đưa ra một truy vấn, Prolog sẽ tìm kiếm trong cơ sở tri thức để xác định xem truy vấn có thể suy luận ra "đúng" hay không.
* Ví dụ: ?- gioi(lan, toan). %Lan có giỏi toán không?

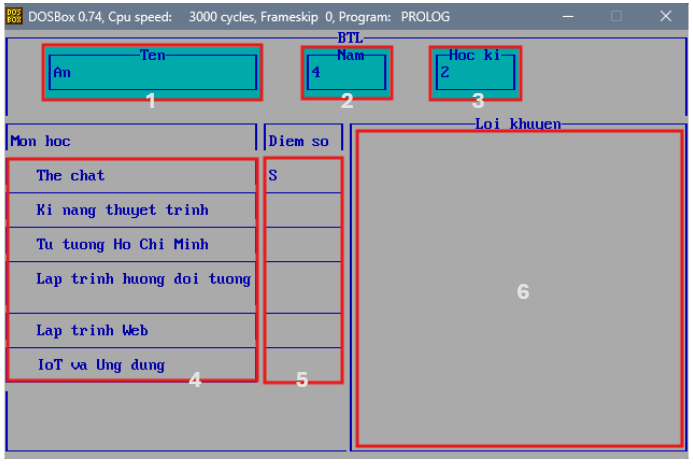
### 3.1.3 Ứng dụng của ngôn ngữ prolog

Ngôn ngữ lập trình prolog có nhiều ứng dụng trong các lĩnh vực khác nhau, đặc biệt là trong trí tuệ nhân tạo và hệ thống chuyên gia. Các ứng dụng phổ biến của prolog:

* Hệ chuyên gia: Prolog rất phù hợp để xây dựng các hệ thống chuyên gia do khả năng xử lý các luật và sự kiện một cách logic. Các hệ thống này có thể đưa ra lời khuyên hoặc dự đoán dựa trên các tập hợp quy tắc và tri thức được cung cấp, như chẩn đoán y khoa, hỗ trợ pháp lý, hay kiểm tra chất lượng sản phẩm.
* Xử lý ngôn ngữ tự nhiên: Với khả năng lập trình dựa trên các quy tắc logic, Prolog được sử dụng để xây dựng các hệ thống xử lý ngôn ngữ tự nhiên, như dịch máy, phân tích cú pháp ngôn ngữ, và chatbot. Khả năng suy diễn của Prolog giúp phân tích và hiểu các cấu trúc ngữ pháp phức tạp trong ngôn ngữ tự nhiên.
* Hệ thống quản lý cơ sở dữ liệu: Prolog có thể được sử dụng để phát triển các hệ thống quản lý cơ sở dữ liệu động. Với cơ chế suy diễn và truy vấn linh hoạt, Prolog có thể xử lý dữ liệu quan hệ phức tạp và thực hiện các thao tác tìm kiếm, lọc dữ liệu một cách hiệu quả.
* Giải quyết bài toán logic và chứng minh định lý: Prolog hỗ trợ giải các bài toán logic, chứng minh định lý và kiểm tra tính đúng đắn của các lý thuyết khoa học. Ngôn ngữ này có thể áp dụng cho các bài toán đòi hỏi suy luận phức tạp và đưa ra các giải pháp hoặc chứng minh các định lý toán học.
* Lập kế hoạch và tối ưu hóa: Khả năng suy diễn của Prolog giúp giải quyết các bài toán lập kế hoạch, lịch trình, và tối ưu hóa, như lập kế hoạch sản xuất, sắp xếp thời khóa biểu, và phân phối nguồn lực.

## 3.2 Giao diện chương trình

### 3.2.1 Giao diện chính

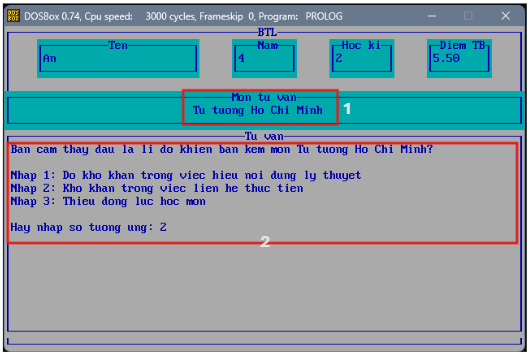


Hình 3.1. Giao diện chính của hệ thống

Giao diện gồm các thành phần chính

1. Trường để nhập tên sinh viên
2. Trường để nhập năm học (1 – 4)
3. Trường để nhập học kì (1 – 2)
4. Danh sách môn học trong học kì
5. Điểm số của từng môn học trong học kì
6. Hiển thị lời khuyên do hệ thống đưa ra

### Giao diện tư vấn cụ thể từng môn học



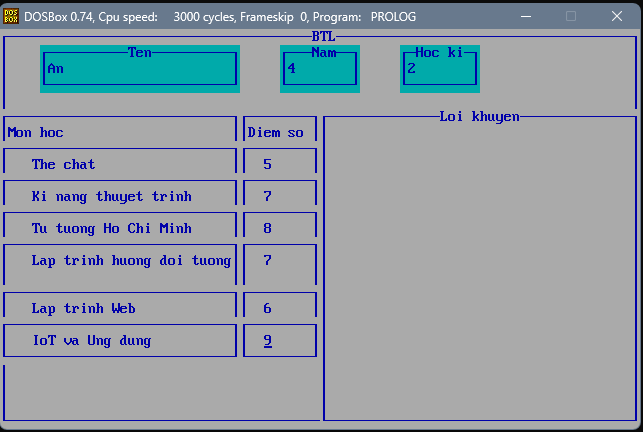
Hình 3.2. Giao diện tư vấn cụ thể

Giao diện gồm các thành phần chính

1. Hiển thị tên môn học người dùng chọn để được tư vấn chi tiết
2. Khung tư vấn để hệ thống đưa ra các câu hỏi, người dùng tương tác với hệ thống bằng cách nhập lựa chọn của mình.

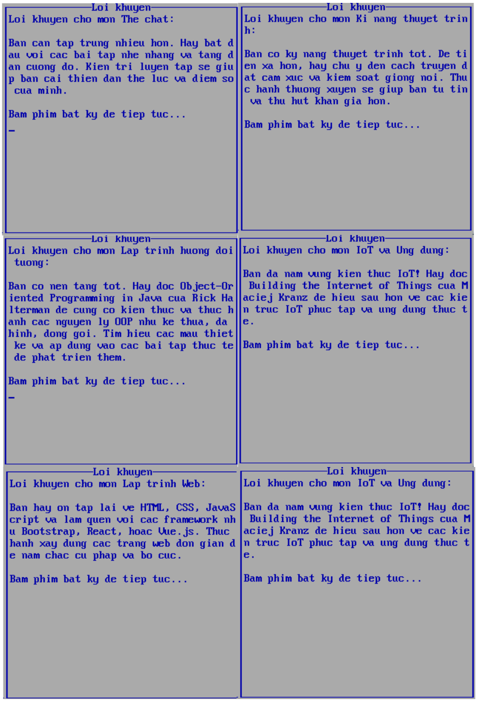
## 3.4 Hỏi đáp

Nhập tên, năm học, kì học và điểm số của từng môn học trong kì học



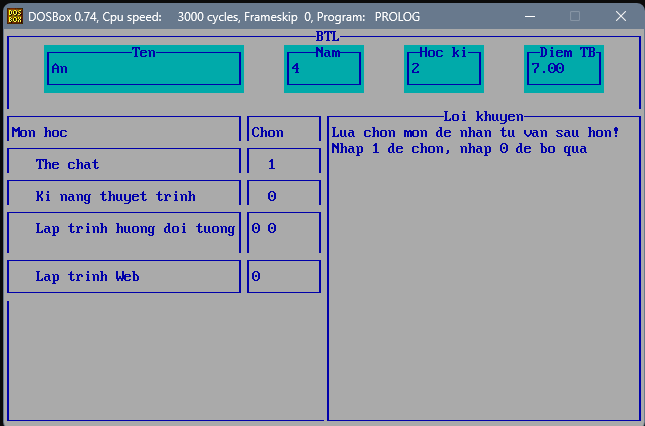
Hình 3.4. Nhập tên, năm học, kì học và điểm số các môn học

Hệ thống đưa ra lời khuyên chung cho từng môn học



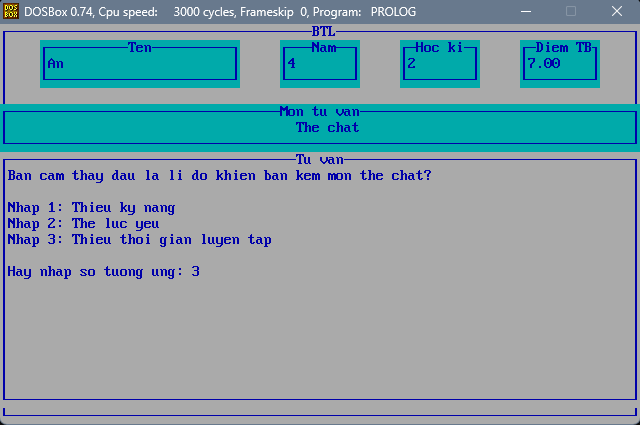
Hình 3.5. Lời khuyên cho từng môn học

Người dùng lựa chọn các môn học cần được tư vấn sâu hơn

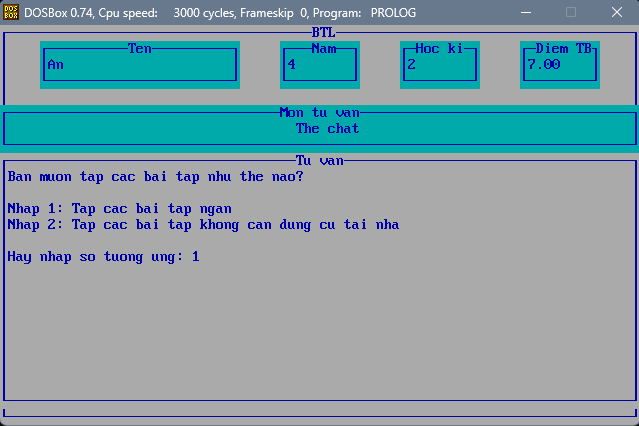


Hình 3.6. Lựa chọn môn học cần tư vấn sâu

Hệ thống hiện ra các câu hỏi về môn học, người dùng trả lời để được tư vấn

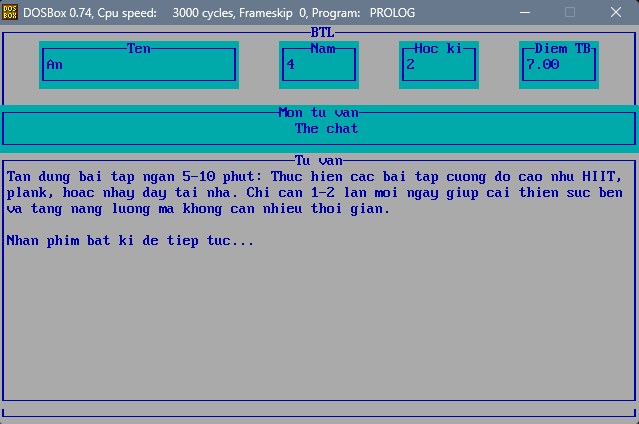


Hình 3.7. Lựa chọn các gợi ý do hệ thống đưa ra

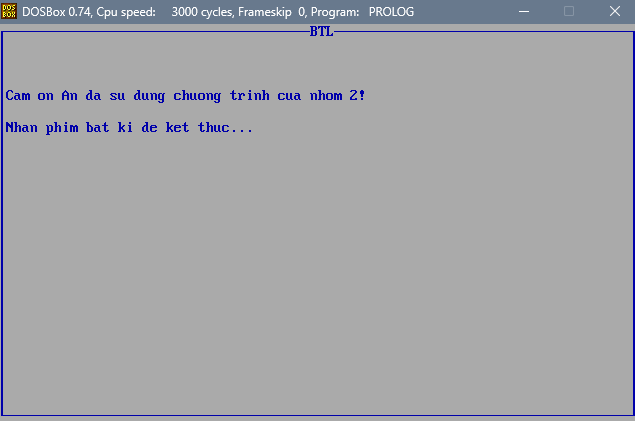


Hình 3.8. Tiếp tục lựa chọn các gợi ý được đưa ra

Dựa trên các lựa chọn, hệ thống đưa ra lời khuyên cụ thể cho người dùng



Hình 3.9. Hệ thống đưa ra lời khuyên cho người dùng



Hình 3.10. Kết thúc chương trình

## 3.4 Kết luận

Chương 3 tập trung vào việc giới thiệu ngôn ngữ lập trình Prolog, xây dựng giao diện và cơ chế tư vấn môn học bằng Prolog, cùng với việc tạo ra chức năng hỏi đáp với chương trình. Prolog, với khả năng xử lý các quy tắc logic và phép suy luận, là lựa chọn lý tưởng cho hệ chuyên gia. Giao diện đã giúp tạo ra trải nghiệm tương tác thân thiện, cho phép người dùng dễ dàng nhập dữ liệu và nhận phản hồi. Chức năng hỏi đáp cho phép người dùng tra cứu thông tin theo nhu cầu, giúp sinh viên có hướng dẫn học tập cá nhân hóa và hiệu quả. Nhờ vào Prolog và thiết kế chương trình, hệ thống tư vấn môn học trở thành một công cụ giúp hỗ trợ quá trình học tập của người dùng một cách rõ ràng và dễ hiểu.

# Kết luận

## 1. Kết quả đạt được

* Hiểu về trí tuệ nhân tạo và hệ chuyên gia: làm rõ về các khái niệm liên quan đến trí tuệ nhân tạo và hệ chuyên gia, giúp xác định được cách ứng dụng của các công nghệ trong các bài toán thực tế.
* Xây dựng hệ chuyên gia dựa trên luật: phát triển một hệ chuyên gia sử dụng luật để tư vấn cho sinh viên về các môn học dựa trên điểm số, kỳ học, và các yếu tố học thuật khác.
* Ứng dụng Prolog trong lập trình hệ chuyên gia: Sử dụng ngôn ngữ Prolog, chúng tôi đã thiết lập các quy tắc và cơ chế suy luận của hệ thống, đồng thời tạo ra giao diện thân thiện để người dùng có thể dễ dàng nhập dữ liệu và nhận kết quả.
* Phát triển giao diện và cơ chế hỏi đáp: Giao diện của chương trình cho phép người dùng nhập thông tin cá nhân và điểm số các môn học, đồng thời sử dụng tính năng hỏi đáp để hỗ trợ sinh viên tìm kiếm thông tin và nhận được các lời khuyên học tập cụ thể.

## 2. Phương hướng phát triển đề tài

* Mở rộng tập luật: Tiếp tục bổ sung và hoàn thiện các luật mới để hỗ trợ nhiều trường hợp học tập khác nhau, bao gồm các môn học còn lại trong chương trình học.
* Nâng cấp giao diện và trải nghiệm người dùng: Thiết kế giao diện trực quan và linh hoạt hơn, cải thiện khả năng tương tác và đáp ứng các câu hỏi phức tạp của người dùng.
* Kết hợp các kỹ thuật AI khác: Ngoài cây quyết định, có thể tích hợp thêm các mô hình học máy để tăng độ chính xác trong tư vấn và dự đoán các yếu tố thành công trong học tập.
* Xây dựng cơ sở dữ liệu phong phú hơn: Tạo cơ sở dữ liệu lưu trữ lịch sử học tập của sinh viên, giúp hệ thống đưa ra những tư vấn cá nhân hóa hơn theo thời gian.

# Chương 4. Phân cụm sinh viên theo điểm bằng thuật toán K-means

## 4.1. Đặt bài toán

Mục tiêu của bài toán là phân cụm sinh viên dựa trên điểm số. Dựa trên kết quả phân cụm, có thể phân tích các đặc điểm nổi bật của từng cụm để hỗ trợ việc định hướng và cải thiện chất lượng học tập của sinh viên.

## 4.2. Phân tích khám phá dữ liệu

### 4.2.1 Khái niệm và mục đích sử dụng

Phân tích khám phá dữ liệu (EDA) là bước đầu tiên trong quy trình phân tích dữ liệu, tập trung vào việc khám phá và hiểu dữ liệu một cách toàn diện. Mục tiêu chính của nó là nhận diện các mẫu, xu hướng, và các vấn đề bất thường như dữ liệu lỗi hoặc giá trị ngoại lai.

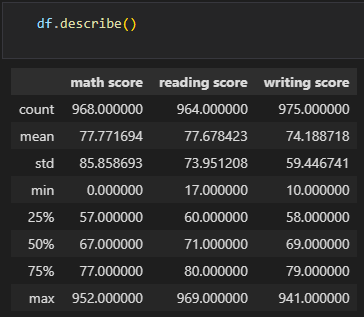
EDA giúp cung cấp một cái nhìn trực quan và định lượng về dữ liệu thông qua các công cụ thống kê mô tả (như giá trị trung bình, độ lệch chuẩn) và trực quan hóa (biểu đồ phân tán, histogram, ...).

Ngoài ra, EDA giúp kiểm tra mối quan hệ giữa các biến, xác định sự tương quan và tìm hiểu đặc điểm phân phối dữ liệu. Đây là bước nền tảng quan trọng để chuẩn bị dữ liệu cho các bước phân tích sâu hơn, như xây dựng mô hình hoặc kiểm định giả thuyết.

### 4.2.2 Áp dụng đối với bài toán phân cụm sinh viên

Kích thước bộ dữ liệu: gồm 1000 dòng và 8 cột

Thống kê cơ bản về dữ liệu



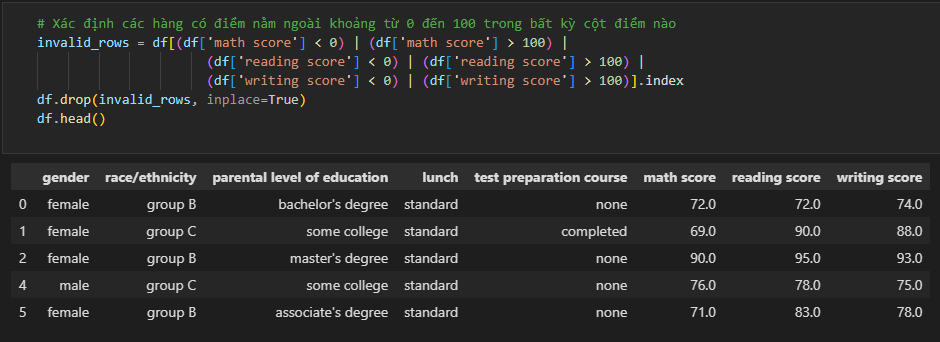
Hình 4.1 Thống kê về bộ dữ liệu

* count: số lượng giá trị không rỗng (non-null) trong mỗi cột.
* mean: giá trị trung bình của từng cột.
* std: độ lệch chuẩn, đo lường sự phân tán của dữ liệu.
* min và max: Giá trị nhỏ nhất và lớn nhất.
* 25%, 50%, 75%: Các percentiles (phân vị), giúp đánh giá phân bố dữ liệu.

### 4.2.3 Tiền xử lí dữ liệu

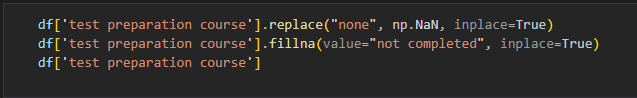
Làm sạch dữ liệu bằng cách xác định, loại bỏ những hàng có giá trị không hợp lệ trong cột điểm (có giá trị < 0 hoặc > 100), đảm bảo dữ liệu hợp lý trước khi phân tích hoặc sử dụng cho các bước tiếp theo.

Kích thước bộ dữ liệu sau khi loại bỏ dữ liệu không hợp lệ: 945

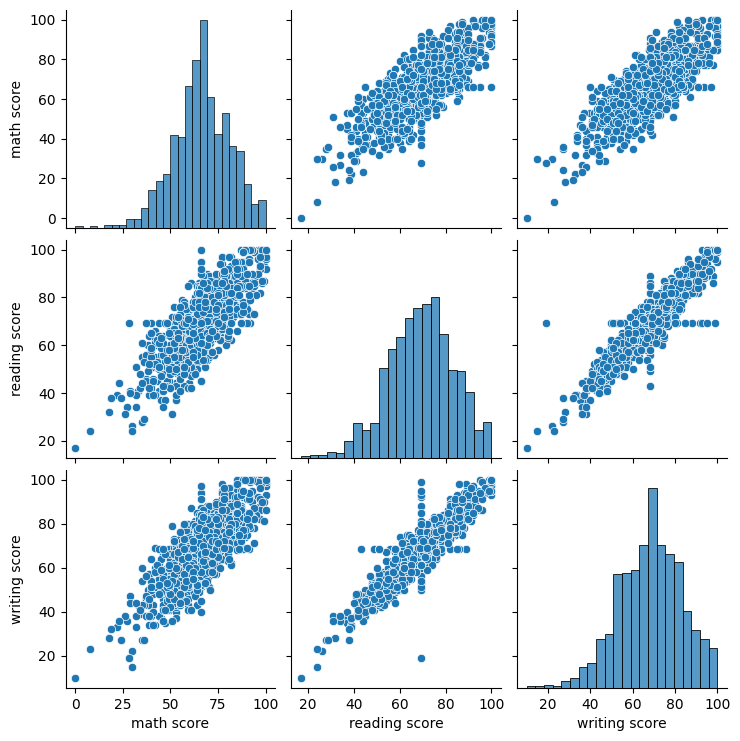


Hình 4.2 Loại bỏ các giá trị không hợp lệ

Thực hiện xử lý giá trị bị thiếu trong các cột math score, reading score, và writing score bằng cách thay thế chúng bằng giá trị trung bình của từng cột. Sau đó, kiểm tra lại để đảm bảo không còn giá trị nào bị thiếu, giúp làm sạch dữ liệu, đảm bảo tính toàn vẹn và sẵn sàng cho các bước tiếp theo.



Hình 4.3 Xử lí giá trị bị thiếu



Hình 4.4 Biểu diễn dữ liệu dưới dạng biểu đồ

## 4.3. Phân cụm sinh viên bằng thuật toán K-means

### 4.3.1 Thuật toán K-means

Thuật toán K-Means là một trong những thuật toán học máy không giám sát kinh điển để giải quyết bài toán phân cụm. Thuật toán này được giới thiệu vào những năm 50 của thế kỷ 20 và là một trong 10 thuật toán được sử dụng nhiều nhất trong lĩnh vực khai phá dữ liệu và phát hiện tri thức.

Trong thuật toán K-means, chúng ta không biết nhãn của từng điểm dữ liệu. Mục đích là làm thể nào để phân dữ liệu thành các cụm khác nhau sao cho dữ liệu trong cùng một cụm có tính chất giống nhau.

K-Means hoạt động bằng cách tìm các cụm dữ liệu sao cho tổng bình phương khoảng cách giữa các điểm dữ liệu và tâm cụm của chúng là nhỏ nhất. Mỗi cụm được xác định bởi một tâm cụm, và mỗi điểm dữ liệu được gán vào cụm gần nó nhất dựa trên khoảng cách Euclide.

Ưu điểm:

* Đơn giản, dễ hiểu và tốc độ nhanh.
* Hiệu quả cho các bài toán với dữ liệu lớn.
* Ứng dụng trong nhiều lĩnh vực: xử lí ảnh, phân tích dữ liệu lớn,…

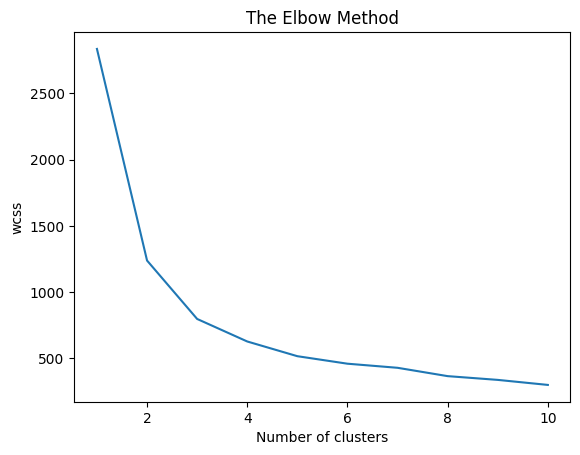
Nhược điểm:

* Cần biết trước số cụm.
* Nhạy cảm với giá trị khởi tạo và các điểm dữ liệu nhiễu.
* Không phù hợp cho các cụm có hình dạng không phải hình cầu.

### 4.3.2 Xác định số cụm dựa trên phương pháp khuỷu tay

Các bước thực hiện:

* Chạy K-Means với nhiều giá trị K khác nhau.
* Tính tổng bình phương khoảng cách trong cụm (WCSS – Within-Cluster Sum of Squares) cho mỗi K.
* Vẽ đồ thị WCSS theo K và tìm “khuỷu tay” – điểm mà việc tăng K không làm giảm WCSS đáng kể.



Hình 4.5 Đồ thị hàm biến dạng của thuật toán K-Means

Trục tung là giá trị của hàm biến dạng và trục hoành là giá trị của số lượng cụm cần phân chia trong thuật toán k-Means.

Điểm khuỷ tay là điểm mà ở đó tốc độ suy giảm của hàm biến dạng sẽ thay đổi nhiều nhất. Tức là kể từ sau vị trí này thì gia tăng thêm số lượng cụm cũng không giúp hàm biến dạng giảm đáng kể. Nếu thuật toán phân chia theo số lượng cụm tại vị trí này sẽ đạt được tính chất phân cụm một cách tổng quát nhất. Trong hình trên thì ta thấy vị trí của điểm khuỷ tay chính là k = 3 vì khi số lượng cụm lớn hơn 2 thì tốc độ suy giảm của hàm biến dạng dường như không đáng kể so với trước đó.

Phương pháp Elbow là một phương pháp thường được sử dụng để lựa chọn số lượng cụm phân chia hợp lý dựa trên biểu đồ, tuy nhiên có một số trường hợp chúng ta sẽ không dễ dàng phát hiện vị trí của Elbow, đặc biệt là đối với những bộ dữ liệu mà qui luật phân cụm không thực sự dễ dàng được phát hiện. Nhưng nhìn chung thì phương pháp Elbow vẫn là một phương pháp tốt nhất được ứng dụng trong việc tìm kiếm số lượng cụm cần phân chia.

### 4.3.3 Phân cụm sinh viên sử dụng thuật toán K-means

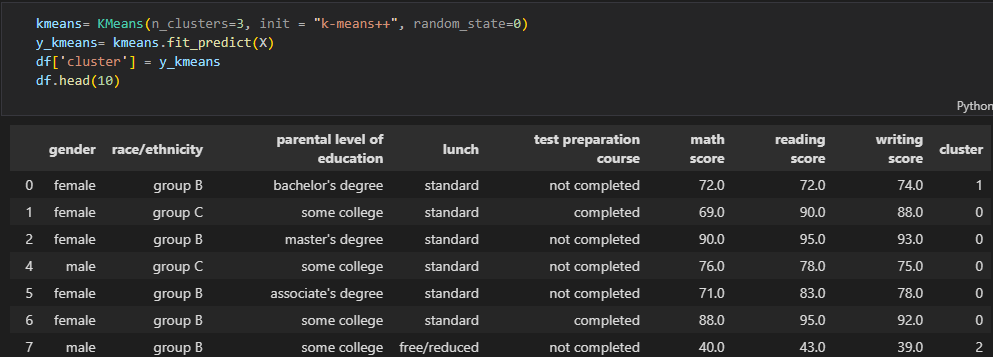
Các bước thực hiện:

Bước 1: Bắt đầu bằng việc chọn k số điểm dữ liệu ngẫu nhiên (cụm) trong tập dữ liệu. K là số cụm cần phân loại, được lựa chọn trước khi thiết lập thuật toán, với bài toán phân cụm sinh viên ta chọn k = 3.

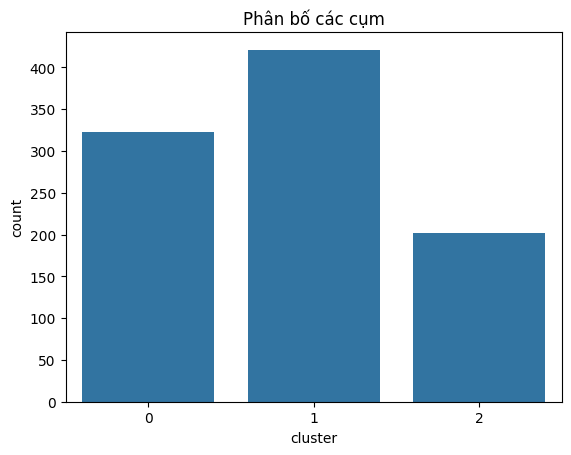
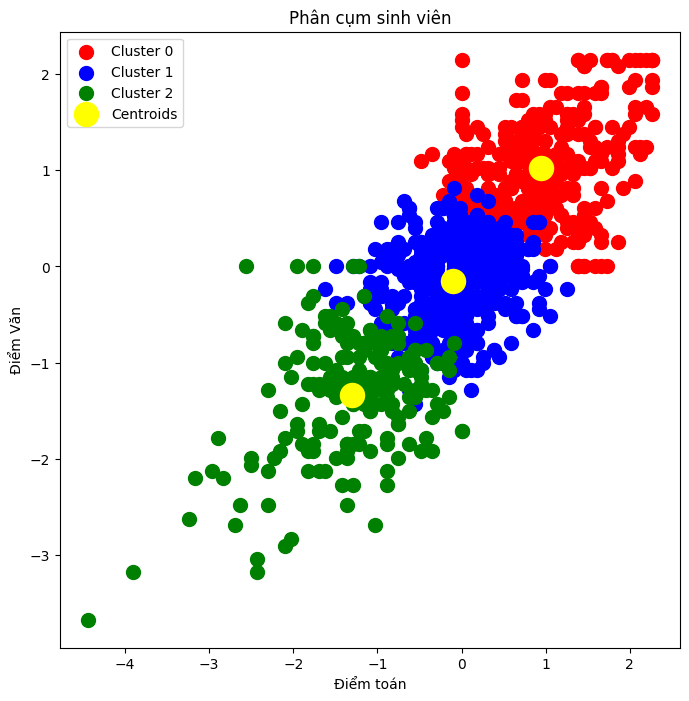
Bước 2: Sau khi có k cụm ban đầu, chúng ta sẽ tính toán khoảng cách giữa từng điểm dữ liệu với k cụm này và gán điểm dữ liệu đó vào cụm gần nó nhất. Khoảng cách giữa hai điểm dữ liệu thường được tính bằng khoảng cách Euclide.

Bước 3: Sau khi đã gán nhãn cho tất cả các điểm dữ liệu, chúng ta cần xác định lại tâm của các cụm để cải thiện hiệu quả của thuật toán. Tâm mới của cụm sẽ được xác định bằng cách tính trung bình vị trí của tất cả các điểm dữ liệu thuộc cụm đó.

Bước 4: Quá trình gán nhãn và cập nhật tâm cụm sẽ được lặp lại cho đến khi tâm cụm không thay đổi sau mỗi vòng lặp (hay chênh lệch đủ nhỏ) hoặc đạt số lần lặp tối đa.



Hình 4.6 Tiến hành phân cụm với k = 3



Hình 4.7 Biểu diễn các cụm trên đồ thị

## 4.4. Kết luận

Trong phần này, nhóm đã trình bày chi tiết phương pháp Elbow để xác định số lượng cụm tối ưu trong thuật toán K-Means. Phương pháp này dựa trên việc phân tích mối quan hệ giữa số lượng cụm và giá trị của hàm biến dạng WCSS

Trong ví dụ phân cụm sinh viên, k = 3 được chọn làm số cụm tối ưu vì tại đó tốc độ suy giảm của hàm biến dạng giảm đáng kể, giúp phân cụm dữ liệu một cách tổng quát và hiệu quả.

Quy trình phân cụm sử dụng thuật toán K-Means được thực hiện qua các bước gán nhãn điểm dữ liệu, cập nhật tâm cụm và lặp lại đến khi hội tụ. Kết quả cuối cùng giúp chia các sinh viên thành 3 nhóm có đặc điểm tương đồng, hỗ trợ việc phân tích và ra quyết định trong các ứng dụng thực tiễn.

Mặc dù phương pháp Elbow đơn giản và phổ biến, việc xác định "điểm khuỷu tay" có thể gặp khó khăn đối với những tập dữ liệu phức tạp hoặc không rõ ràng. Tuy nhiên, phương pháp này vẫn là một công cụ hiệu quả trong việc lựa chọn số cụm cần thiết cho phân tích dữ liệu.

# Tài liệu tham khảo

1. Turbo prolog 2.0 reference guide 2.0, Borderland International Inc, 1988
2. Prolog programing for Artificial Intelligence 4th, Ivan Bratko, Addison Wesley, 2011
3. Expert Systems Prolog, Dennis Merrit, Independently published, 2017