**Đại Học Kinh Tế Quốc Dân**

**Trường Công Nghệ**

**Khoa Công Nghệ Thông Tin**

**----\*\*\*----**



**Báo Cáo Tiểu Luận**

**Đề Tài: Xây Dựng Hệ Thống Chuẩn Đoán Bệnh**

**Trong Lĩnh Vực Y Học Bằng Học Máy**

|  |  |
| --- | --- |
| Giảng viên | : TS. Lưu Minh Tuấn |
| Họ và tên sinh viên | : Nguyễn Trọng Vỹ |
| Mã sinh viên | : 11227025 |
| Lớp  Lớp Học Phần | : Công Nghệ Thông Tin 64B  : Trí tuệ nhân tạo 02 |

**Hà Nội - 2025**

**Lời Nói Đầu**

Trong thời đại công nghệ 4.0, trí tuệ nhân tạo (AI) đang dần khẳng định vai trò quan trọng trong nhiều lĩnh vực, đặc biệt là trong ngành y tế. Việc ứng dụng các kỹ thuật học máy (Machine Learning) giúp nâng cao hiệu quả trong việc hỗ trợ chẩn đoán bệnh, rút ngắn thời gian và giảm thiểu sai sót trong quá trình đánh giá tình trạng sức khỏe của bệnh nhân.

Nhằm tiếp cận với những ứng dụng thực tiễn của trí tuệ nhân tạo, em đã lựa chọn đề tài **“Hệ thống chẩn đoán bệnh trong lĩnh vực ý học bằng học máy”**. Đề tài hướng đến việc xây dựng một mô hình đơn giản có khả năng phân tích các triệu chứng cơ bản do người dùng nhập vào để đưa ra dự đoán sơ bộ về bệnh lý có thể mắc phải. Đây là một bước khởi đầu để hình dung cách AI có thể hỗ trợ bác sĩ và người dùng trong quá trình sàng lọc ban đầu.

Trong quá trình thực hiện đề tài, em đã áp dụng kiến thức về học máy, tiền xử lý dữ liệu, xây dựng mô hình, cũng như triển khai giao diện sử dụng trên môi trường website nhằm tăng tính trực quan và tiện dụng.

Em hy vọng sản phẩm này có thể minh họa được tiềm năng của trí tuệ nhân tạo trong y học, đồng thời giúp em củng cố kiến thức và kỹ năng lập trình thực tế.

**Chương 1: Giới thiệu đề tài**

* 1. **Phát biểu đề tài**

Trong những năm gần đây, trí tuệ nhân tạo (AI) đã trở thành một trong những lĩnh vực công nghệ mũi nhọn, được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau như tài chính, giáo dục, nông nghiệp và đặc biệt là y tế. Trong lĩnh vực chăm sóc sức khỏe, AI và các thuật toán học máy (Machine Learning) đang góp phần thay đổi cách con người phát hiện, chẩn đoán và điều trị bệnh.

Với mục tiêu tìm hiểu và ứng dụng các kiến thức đã học về trí tuệ nhân tạo và học máy, em quyết định lựa chọn đề tài **" Hệ thống chẩn đoán bệnh trong lĩnh vực ý học bằng học máy "**. Đề tài tập trung vào việc xây dựng một mô hình học máy có khả năng phân tích các triệu chứng đầu vào từ người dùng và dự đoán loại bệnh tương ứng dựa trên dữ liệu huấn luyện có sẵn.

Hệ thống được xây dựng nhằm hỗ trợ người dùng trong việc nhận diện sớm một số triệu chứng cơ bản và hướng đến việc phát triển một nền tảng đơn giản, dễ sử dụng, triển khai trên nền tảng web. Đề tài mang tính thực tiễn cao, giúp em rèn luyện kỹ năng thu thập dữ liệu, xây dựng mô hình học máy, lập trình giao diện và triển khai sản phẩm thực tế.

* 1. **Mục tiêu của đề tài**

Mục tiêu chính của đề tài là xây dựng một hệ thống hỗ trợ chẩn đoán bệnh đơn giản, dựa trên các biểu hiện triệu chứng mà người dùng nhập vào, từ đó đưa ra dự đoán về loại bệnh có khả năng xảy ra. Cụ thể, đề tài hướng tới các mục tiêu sau:

* Ứng dụng mô hình học máy để phân tích dữ liệu triệu chứng và dự đoán bệnh.
* Thiết kế giao diện người dùng thân thiện, trực quan, giúp người dùng dễ dàng nhập dữ liệu và nhận kết quả dự đoán.
* Huấn luyện và đánh giá mô hình dựa trên tập dữ liệu bệnh phổ biến, đảm bảo mô hình có độ chính xác tương đối.
* Triển khai hệ thống hoàn chỉnh, có thể chạy trình duyệt web.
* Tăng cường hiểu biết và kỹ năng thực hành về trí tuệ nhân tạo, học máy, xử lý dữ liệu và phát triển phần mềm thực tế cho sinh viên.
  1. **Phạm vi của đề tài**

Do thời gian và nguồn lực có hạn, đề tài được thực hiện trong phạm vi giới hạn sau:

* Phạm vi dữ liệu: Dữ liệu đầu vào bao gồm một số triệu chứng cơ bản như sốt, ho, đau đầu, mệt mỏi, đau nhức cơ thể,... Những triệu chứng này được dùng để chẩn đoán một số bệnh thông thường như cảm lạnh, sốt xuất huyết, sốt rét, hen suyễn,…
* Phạm vi mô hình: Đề tài sử dụng mô hình học máy cơ bản (Random Forest) không tập trung vào các mô hình phức tạp như mạng nơ-ron sâu (Deep Learning).
* Phạm vi kỹ thuật: Hệ thống được xây dựng bằng ngôn ngữ Python, sử dụng các thư viện như Scikit-learn để huấn luyện mô hình, Flask hoặc để xây dựng giao diện người dùng.
* Phạm vi triển khai: Hệ thống được triển khai dưới dạng và web cơ bản, phục vụ nhu cầu sử dụng thử nghiệm và học tập, không thay thế chẩn đoán y tế chuyên môn.
* Giới hạn của đề tài: Hệ thống chỉ mang tính chất hỗ trợ, không thể thay thế chẩn đoán từ bác sĩ. Kết quả dự đoán chỉ mang tính tham khảo.
  1. **Đối tượng nghiên cứu của đề tài**

Đề tài tập trung nghiên cứu các đối tượng sau:

* Các triệu chứng bệnh lý thường gặp: Bao gồm các biểu hiện cơ bản như sốt, ho, đau đầu, mệt mỏi, đau nhức cơ thể,… Đây là những dấu hiệu phổ biến có thể liên quan đến nhiều loại bệnh thông thường.
* Các bệnh lý phổ biến: Đề tài tập trung vào việc dự đoán một số loại bệnh thường gặp như:
  + Cảm lạnh thông thường
  + Sốt xuất huyết
  + Sốt rét
  + Hen suyễn
  + Sốt thông thường
  + Sổ mũi
  + Đau nhức cơ thể
* Mô hình học máy trong phân loại bệnh: Nghiên cứu và ứng dụng mô hình học máy Random Forest để thực hiện bài toán phân loại.
* Công nghệ triển khai phần mềm: Đối tượng nghiên cứu còn bao gồm các công cụ, thư viện phục vụ triển khai hệ thống như:
  + Scikit-learn để huấn luyện và lưu trữ mô hình
  + Pandas để đọc dữ liệu huấn luyện
  + Flask để xây dựng giao diện
  + Joblib để nạp mô hình đã huấn luyện
  1. **Ý nghĩa khoa học và thực tiễn**
* Ý nghĩa khoa học:
  + Ứng dụng học máy trong y học: Việc áp dụng học máy vào lĩnh vực y học giúp mở rộng và cải tiến các phương pháp phân tích, từ đó đóng góp vào sự phát triển của ngành khoa học dữ liệu và học máy. Mô hình học máy có khả năng phân tích và rút ra các mẫu từ các dữ liệu y tế lớn, hỗ trợ nâng cao hiệu quả chẩn đoán bệnh.
  + Khám phá mối liên hệ mới: Các mô hình học máy có thể tìm ra những mối quan hệ chưa được khám phá giữa các yếu tố trong dữ liệu bệnh lý (như triệu chứng, dấu hiệu lâm sàng, kết quả xét nghiệm) mà các phương pháp truyền thống không thể phát hiện.
  + Nâng cao khả năng dự đoán: Mô hình học máy có thể giúp cải thiện độ chính xác và hiệu suất của các hệ thống chẩn đoán, đặc biệt trong các bệnh lý phức tạp hoặc ít được nghiên cứu.
* Ý nghĩa thực tiễn:
  + Hỗ trợ chẩn đoán bệnh nhanh chóng và chính xác: Các mô hình học máy có thể giúp các bác sĩ và nhân viên y tế đưa ra quyết định chẩn đoán một cách nhanh chóng, chính xác hơn, giảm thiểu sai sót và tiết kiệm thời gian. Điều này đặc biệt quan trọng trong các tình huống khẩn cấp, nơi mà việc chẩn đoán kịp thời có thể cứu sống bệnh nhân.
  + Giảm chi phí y tế: Việc tự động hóa quá trình chẩn đoán bệnh giúp giảm chi phí cho việc khám chữa bệnh, nhờ vào việc sử dụng công nghệ để phân tích và phát hiện bệnh sớm. Điều này có thể giúp giảm gánh nặng cho hệ thống y tế.
  + Ứng dụng rộng rãi trong chăm sóc sức khỏe cộng đồng: Mô hình học máy có thể được triển khai trong các hệ thống y tế ở các vùng sâu, vùng xa, nơi có ít bác sĩ và thiết bị y tế. Nhờ đó, những người dân ở khu vực này cũng có thể được chẩn đoán và điều trị sớm.

**Chương 2: Cơ Sở lý thuyết**

* 1. **Giới thiệu về học máy**

**Học máy** (Machine Learning) là một nhánh con của trí tuệ nhân tạo (AI), nghiên cứu và phát triển các thuật toán cho phép máy tính học hỏi từ dữ liệu và cải thiện khả năng thực hiện các tác vụ mà không cần lập trình trực tiếp. Học máy được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, từ nhận diện hình ảnh, nhận dạng giọng nói, phân tích dữ liệu, đến y học và tài chính.

* + 1. **Định nghĩa học máy**

Học máy là quá trình mà máy tính tự học và cải thiện hiệu suất của mình qua kinh nghiệm, thông qua việc phân tích và nhận diện các mẫu (patterns) trong dữ liệu mà không cần được lập trình một cách cụ thể cho từng trường hợp.

* + 1. **Các loại học máy**

Học máy có thể được chia thành ba loại chính:

* **Học có giám sát (Supervised Learning)**:
  + Trong học có giám sát, mô hình học từ một tập dữ liệu đã được gắn nhãn, tức là mỗi ví dụ trong dữ liệu đã có sẵn kết quả (output) chính xác. Mục tiêu của mô hình là học được mối quan hệ giữa đầu vào và đầu ra để có thể dự đoán kết quả cho những ví dụ chưa được gắn nhãn.
  + Ví dụ: Dự đoán bệnh dựa trên các thông số sức khỏe như huyết áp, đường huyết.
* **Học không giám sát (Unsupervised Learning)**:
  + Trong học không giám sát, mô hình học từ dữ liệu chưa có nhãn, tức là không có kết quả chính xác được gắn kèm với các ví dụ. Mục tiêu là tìm ra cấu trúc hoặc mẫu trong dữ liệu mà không cần biết trước kết quả.
  + Ví dụ: Phân nhóm bệnh nhân thành các nhóm có triệu chứng giống nhau.
* **Học tăng cường (Reinforcement Learning)**:
  + Học tăng cường là một hình thức học mà trong đó một tác nhân (agent) học cách tối ưu hóa hành động của mình thông qua việc nhận phản hồi từ môi trường. Mô hình sẽ nhận được các phần thưởng hoặc hình phạt dựa trên các hành động mà nó thực hiện, và qua thời gian sẽ học được cách đưa ra quyết định tốt nhất.
  + Ví dụ: Hệ thống robot học cách đi lại qua môi trường mà không va phải vật cản.
    1. **Các thuật toán học máy phổ biến**
* **Hồi quy (Regression):** Được sử dụng trong các bài toán dự đoán giá trị liên tục, như dự đoán giá nhà, nhiệt độ, hoặc tình trạng sức khỏe. Các thuật toán hồi quy phổ biến bao gồm hồi quy tuyến tính và hồi quy logistic.
* **Phân loại (Classification):** Được sử dụng trong các bài toán phân nhóm đối tượng vào các lớp khác nhau. Ví dụ: Phân loại bệnh nhân có hoặc không có bệnh tiểu đường dựa trên các đặc điểm sinh lý học. Các thuật toán phân loại phổ biến bao gồm cây quyết định (Decision Trees), máy vectơ hỗ trợ (SVM), và mạng nơ-ron nhân tạo (ANN).
* **Clustering (Phân nhóm):** Đây là một kỹ thuật học không giám sát, được sử dụng để phân loại các đối tượng thành các nhóm (clusters) dựa trên sự tương đồng. Các thuật toán phân nhóm phổ biến bao gồm K-means và DBSCAN.
* **Mạng nơ-ron nhân tạo (Neural Networks):** Được sử dụng trong các bài toán phức tạp, như nhận diện hình ảnh và giọng nói. Các mạng nơ-ron bao gồm các lớp các tế bào nơ-ron, mô phỏng cách thức hoạt động của bộ não con người.
  + 1. **Quy trình học máy**
    2. **Ứng dụng của học máy trong y học**
    3. **Lợi ích và thách thức của học máy**
  1. **Các mô hình phân loại**
  2. **Mô hình sử dụng**
  3. **Các thư viện hỗ trợ**

**Chương 3: Xây dựng hệ thống phần mềm**

* 1. Thiết kế hệ thống
  2. Cài đặt và thử nghiệm
  3. Giao diện sử dụng

**Chương 4: Đánh giá và kết luận**

* 1. Kết quả dự đoán
  2. Các hạn chế còn tồn tại
  3. Hướng cải tiến