**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN**

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

**Logo

Description automatically generated**

**TIỂU LUẬN**

**Đề tài:** “Nghiên cứu về học máy (machine learning) và cài đặt ứng dụng thử nghiệm mô hình Naïve Bayes”

*Sinh viên thực hiện :* Đỗ Tuấn Đức

*Lớp :* Công nghệ thông tin 62A

*Mã sinh viên :* 11200832

*Bộ môn :* Trí tuệ nhân tạo

*Giảng viên :* Lưu Minh Tuấn

*Hà Nội, 2022*

**MỤC LỤC**

*Đặt vấn đề 3*

*PHẦN I: GIỚI THIỆU CHUNG VỀ TRÍ TUỆ NHÂN TẠO. 4*

*I.1. Trí tuệ nhân tạo là gì. 4*

*I.2. Mục tiêu của trí tuệ nhân tạo. 5*

*I.3. Ứng dụng AI vào đời sống. 6*

*I.4. Một số mặt tiêu cực của trí tuệ nhân tạo. 7*

*PHẦN II: TÌM HIỂU VỀ MÔ HÌNH PHÂN LỚP NAVIE BAYES. 7*

*II.1. Định lý Bayes... 7*

*II.2. Giả định Naïve. 8*

*II.3. Bộ phân loại Naive Bayes. 9*

*II.4. Các phân phối thường dung cho P(xi|c). 10*

*PHẦN III: TÌM HIỂU VỀ COVID 19. 10*

*III.1. Tìm hiểu chung về covid 19. 10*

*III.2. Cách thức lây lan và triệu chứng thường gặp. 11*

*III.3. Các biểu hiện của người bị dương tính với Covid 19. 13*

*III.4. Dữ liệu của các bệnh nhân nhiễm Covid 19 14*

*IV. ỨNG DỤNG MÔ HÌNH PHÂN LỚP NAÏVE BAYES TRONG CHUẨN ĐOÁN NGƯỜI MẮC COVID 19 THÔNG QUA CÁC TRIỆU CHỨNG VÀ BIỂU HIỆN.14*

*IV.1. Dữ liệu của các bệnh nhân nhiễm Covid 19. 14*

*IV.2. Cài đặt thủ công. 15*

*IV.3. Các vấn đề gặp phải. 17*

*V.TỔNG KẾT 17*

*Đặt vấn đề*

Ngày nay, thế giới đang phát triển từng ngày, lượng thông tin ngày càng lớn và phức tạp, con người không thể lưu trữ thêm kiến thức mà phải nhờ đến một thứ khác để lưu trữ và giải quyết. Thời đại công nghiệp 4.0 phát triển nhanh chóng, chúng ta đã có giải pháp đó là trí tuệ nhân tạo. Trí tuệ nhân tạo với tên tiếng anh Artificial Intelligence hay còn được viết tắt AI, được định nghĩa một cách đơn giản, dễ hiểu là trí thông minh được thể hiện bằng máy móc do con người tạo ra. Điều này ngược lại với trí thông minh của con người hoặc trí thông minh của một vài loài động vật ngoài tự nhiên khác. Nghiên cứu và phát triển trí tuệ nhân tạo là làm cho những cỗ máy "vô tri, vô giác" có được khả năng của trí tuệ và xử lý thông minh của con người. Ở thời điểm hiện tại, AI đang dần có ảnh hưởng lớn đến cuộc sống của chúng ta. Chúng xuất hiện ở trong mọi lĩnh vực trong đời sống như y tế, ngân hàng, tài chính, trò chơi,… Ngoài ra AI còn giúp chúng ta trong việc đưa ra quyết định, thống kê,…

Tìm hiểu về trí tuệ nhân tạo, chúng ta sẽ có rất nhiều thuật toán khác nhau để phục vụ các mục đích khác nhau. Ở bài tập lớn này, em sẽ tìm hiểu về kĩ thuật t Naïve Bayes. Bộ phân lớp Bayes là một giải thuật thuộc lớp giải thuật thống kê, nó có thể dự đoán xác suất của một phần tử dữ liệu thuộc vào một lớp là bao nhiêu. Phân lớp Bayes được dựa trên định lý Bayes (định lý được đặt theo tên tác giả của nó là Thomas Bayes).

Do quá trình tìm hiểu còn có sai sót, em mong thầy có thể giúp đỡ để có bài tiểu luẩn tốt nhất.

**PHẦN I: GIỚI THIỆU CHUNG VỀ TRÍ TUỆ NHÂN TẠO.**

**I.1. Trí tuệ nhân tạo là gì.**

Trong [khoa học máy tính](https://vi.wikipedia.org/wiki/Khoa_h%E1%BB%8Dc_m%C3%A1y_t%C3%ADnh), trí tuệ nhân tạo hay AI ([tiếng Anh](https://vi.wikipedia.org/wiki/Ti%E1%BA%BFng_Anh): *artificial intelligence*), đôi khi được gọi là trí thông minh nhân tạo, là [trí thông minh](https://vi.wikipedia.org/wiki/Tr%C3%AD_th%C3%B4ng_minh) được thể hiện bằng [máy móc](https://vi.wikipedia.org/wiki/M%C3%A1y_m%C3%B3c), trái ngược với trí thông minh tự nhiên của con người. Thông thường, thuật ngữ "trí tuệ nhân tạo" thường được sử dụng để mô tả các máy móc (hoặc máy tính) có khả năng bắt chước các chức năng "nhận thức" mà con người thường phải liên kết với [tâm trí](https://vi.wikipedia.org/wiki/T%C3%A2m_tr%C3%AD), như "học tập" và "giải quyết vấn đề".

Trí tuệ nhân tạo có thể được phân thành ba loại hệ thống khác nhau:

* Trí tuệ nhân tạo phân tích, lấy cảm hứng từ con người và nhân tạo. AI phân tích chỉ có các đặc điểm phù hợp với [trí tuệ nhận thức](https://vi.wikipedia.org/wiki/Nh%E1%BA%ADn_th%E1%BB%A9c),
* Tạo ra một đại diện nhận thức về thế giới và sử dụng học tập dựa trên kinh nghiệm trong quá khứ để thông báo các quyết định trong tương lai. AI lấy cảm hứng từ con người có các yếu tố từ [trí tuệ](https://vi.wikipedia.org/wiki/Tr%C3%AD_tu%E1%BB%87_x%C3%BAc_c%E1%BA%A3m) nhận thức và [cảm xúc](https://vi.wikipedia.org/wiki/Tr%C3%AD_tu%E1%BB%87_x%C3%BAc_c%E1%BA%A3m);
* Hiểu cảm xúc của con người, ngoài các yếu tố nhận thức và xem xét chúng trong việc [ra quyết định](https://vi.wikipedia.org/w/index.php?title=Ra_quy%E1%BA%BFt_%C4%91%E1%BB%8Bnh&action=edit&redlink=1). AI nhân cách hóa cho thấy các đặc điểm của tất cả các loại năng lực (nghĩa là trí tuệ nhận thức, cảm xúc và xã hội), có khả năng tự ý thức và tự nhận thức được trong các tương tác.

**I.2. Mục tiêu của trí tuệ nhân tạo.**

Các nhà nghiên cứu đầu tiên đã phát triển các thuật toán bắt chước theo lý luận từng bước mà con người sử dụng khi giải quyết các câu đố hoặc đưa ra các phương pháp loại trừ logic. Vào cuối những năm 1980 và 1990, nghiên cứu về AI đã phát triển các phương pháp xử lý thông tin không chắc chắn hoặc không đầy đủ, sử dụng các khái niệm từ [xác suất](https://vi.wikipedia.org/wiki/X%C3%A1c_su%E1%BA%A5t) và [kinh tế](https://vi.wikipedia.org/wiki/Kinh_t%E1%BA%BF).

Đối với những vấn đề khó, các thuật toán bắt buộc phải có phần cứng đủ mạnh để thực hiện phép tính toán khổng lồ - để trải qua "vụ nổ tổ hợp": lượng bộ nhớ và thời gian tính toán có thể trở nên vô tận nếu giải quyết một vấn đề khó. Mức độ ưu tiên cao nhất là tìm kiếm các thuật toán giải quyết vấn đề.

Con người thường sử dụng các phán đoán nhanh và trực quan chứ không phải là phép khấu trừ từng bước mà các nghiên cứu AI ban đầu có thể mô phỏng. AI đã tiến triển bằng cách sử dụng cách giải quyết vấn đề "biểu tượng phụ": cách tiếp cận tác nhân được thể hiện nhấn mạnh tầm quan trọng của các kỹ năng cảm biến động đến lý luận cao hơn; nghiên cứu mạng thần kinh cố gắng để mô phỏng các cấu trúc bên trong não làm phát sinh kỹ năng này. Các phương pháp tiếp cận thống kê đối với AI bắt chước khả năng của con người.

**I.3. Ứng dụng AI vào đời sống.**

*Trong ngành vận tải*

Trí tuệ nhân tạo được ứng dụng trên những phương tiện vận tải tự lái, điển hình là ô tô. Sự ứng dụng này góp phần mang lại lợi ích kinh tế cao hơn nhờ khả năng cắt giảm chi phí cũng như hạn chế những tai nạn nguy hiểm đến tính mạng.

Vào năm 2016, Otto, hãng phát triển xe tự lái thuộc Uber đã vận chuyển thành công 50.000 lon bia Budweisers bằng xe tự lái trên quãng đường dài 193 km. Theo dự đoán của công ty tư vấn công nghệ thông tin Gartner, trong tương lai, những chiếc xe có thể kết nối với nhau thông qua Wifi để đưa ra những lộ trình vận tải tốt nhất.

*Trong sản xuất*

Trí tuệ nhân tạo được ứng dụng để xây dựng những quy trình sản xuất tối ưu hơn. Công nghệ AI có khả năng phân tích cao, làm cơ sở định hướng cho việc ra quyết định trong sản xuất.

*Trong y tế*

Ứng dụng tiêu biểu của trí tuệ nhân tạo trong lĩnh vực y tế là máy bay thiết bị bay không người lái được sử dụng trong những trường hợp cứu hộ khẩn cấp. Thiết bị bay không người lái có tốc độ nhanh hơn xe chuyên dụng đến 40% và vô cùng thích hợp để sử dụng ở những nơi có địa hình hiểm trở. Ngoài ra AI có thể hỗ trợ các nhân viên y tế trong việc hỗ trợ chuẩn đoán bệnh và đưa ra các giải pháp.

*Trong giáo dục*

Sự ra đời của trí tuệ nhân tạo giúp tạo ra những thay đổi lớn trong lĩnh vực giáo dục. Các hoạt động giáo dục như chấm điểm hay dạy kèm học sinh có thể được tự động hóa nhờ công nghệ AI. Nhiều trò chơi, phần mềm giáo dục ra đời đáp ứng nhu cầu cụ thể của từng học sinh, giúp học sinh cải thiện tình hình học tập theo tốc độ riêng của mình.

Trí tuệ nhân tạo còn có thể chỉ ra những vấn đề mà các khóa học cần phải cải thiện. Chẳng hạn như khi nhiều học sinh được phát hiện là gửi đáp án sai cho bài tập, hệ thống sẽ thông báo cho giáo viên đồng thời gửi thông điệp đến học sinh để chỉnh sửa đáp án phù hợp. Công nghệ AI còn có khả năng theo dõi sự tiến bộ của học sinh và thông báo đến giáo viên khi phát hiện ra vấn đề đối với kết quả học tập của học sinh.

Hơn nữa, sinh viên còn có thể học hỏi từ bất cứ nơi nào trên thế giới thông qua việc sử dụng những phần mềm có hỗ trợ AI. Công nghệ AI cũng cung cấp dữ liệu nhằm giúp sinh viên lựa chọn được những khóa học tốt nhất cho mình.

*Trong truyền thông*

Đối với lĩnh vực truyền thông, sự phát triển của trí tuệ nhân tạo góp phần làm thay đổi cách thức tiếp cận đối với khách hàng mục tiêu. Nhờ những ưu điểm của công nghệ AI, các công ty có thể cung cấp quảng cáo vào đúng thời điểm, đúng khách hàng tiềm năng, dựa trên việc phân tích các đặc điểm về nhân khẩu học, thói quen hoạt động trực tuyến và những nội dung mà khách hàng thường xem trên quảng cáo.

*Trong ngành dịch vụ*

Công nghệ AI giúp ngành dịch vụ hoạt động tối ưu hơn và góp phần mang đến những trải nghiệm mới mẻ hơn và tốt hơn cho khách hàng. Thông qua việc thu thập và phân tích dữ liệu, công nghệ AI có thể nắm bắt thông tin về hành vi sử dụng dịch vụ của khách hàng, từ đó mang lại những giải pháp phù hợp với nhu cầu của từng khách hàng.

**I.4. Một số mặt tiêu cực của trí tuệ nhân tạo.**

Cuộc sống không ngừng thay đổi, và một điều tất nhiên rằng những cá nhân, tập thể không thể thay đổi để thích nghi được với nhịp sống mới sẽ bị đào thải. *Lợi ích của trí tuệ nhân tạo*là tạo nhiều cơ hội việc làm, nhưng đồng thời nó cũng sẽ là con dao hai lưỡi tước mất việc làm của các đối tượng lao động học vấn thấp, lao động chân tay hoặc mất việc do các công việc được tự động hóa. Theo thống kê, ⅓ công việc trên 32 quốc gia biến mất do tự động hóa. Đây hiện cũng là một vấn đề đau đầu đối với các nhà lãnh đạo, nhưng họ vẫn hy vọng rằng nhân công sẽ thay đổi và bắt kịp với tốc độ phát triển của AI.  Một số nguy cơ tìm ẩn đằng sau lĩnh vực trí tuệ nhân tạo:

* Nguy cơ chiến tranh xâm lược khi sử dụng AI sai mục đích
* Vũ khí thông minh
* Xâm phạm quyền riêng tư
* Thúc đẩy sự thiên vị xã hội

**PHẦN II: TÌM HIỂU VỀ MÔ HÌNH PHÂN LỚP NAVIE BAYES.**

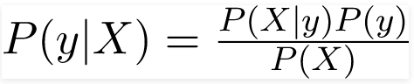
**II.1. Định lý Bayes.**

Text

Description automatically generated Định lý Bayes tìm xác suất của một sự kiện xảy ra với xác suất của một sự kiện khác đã xảy ra. Định lý Bayes được phát biểu về mặt toán học dưới dạng phương trình sau:

trong đó A và B là các sự kiện và P (B) khác 0.

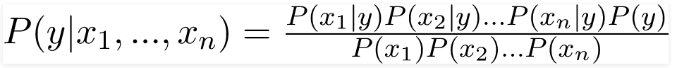
* Về cơ bản, chúng ta đang cố gắng tìm xác suất của sự kiện A, với điều kiện sự kiện B là đúng. Sự kiện B cũng được gọi là bằng chứng.
* P (A) là tiên nghiệm của A (xác suất trước, tức là xác suất của sự kiện trước khi bằng chứng được nhìn thấy). Bằng chứng là một giá trị thuộc tính của một cá thể không xác định (ở đây, nó là sự kiện B).
* P (A|B) là xác suất hậu kỳ của B, tức là xác suất của sự kiện sau khi bằng chứng được nhìn thấy hay có thể hiểu là xác suất xảy ra A khi xảy ra B

Bây giờ, liên quan đến tập dữ liệu của chúng ta, chúng ta có thể áp dụng định lý Bayes theo cách sau:

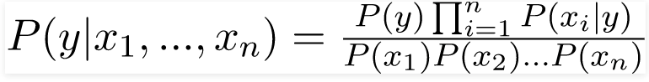
Text

Description automatically generatedtrong đó, y là biến lớp và X là vectơ đặc trưng phụ thuộc (có kích thước n) trong đó:

**II.2. Giả định Naïve.**

 Đặt một giả định Naive cho định lý Bayes, tức là sự độc lập giữa các đối tượng địa lý. Vì vậy, bây giờ, chúng ta chia bằng chứng thành các phần độc lập. Nếu hai sự kiện A và B bất kỳ là độc lập, thì P(A, B) = P(A).P(B)

Do đó, chúng ta đạt được kết quả:

hay

Chúng ta cần tạo một mô hình phân loại. Đối với điều này, chúng ta tìm xác suất của bộ đầu vào đã cho cho tất cả các giá trị có thể có của biến lớp y và chọn đầu ra với xác suất tối đa. Điều này có thể được biểu thị bằng toán học như sau:



Vì vậy, cuối cùng, chúng ta chỉ còn lại nhiệm vụ tính P (y) và P (xi | y).

*Lưu ý rằng P (y) còn được gọi là xác suất lớp và P (xi | y) được gọi là xác suất có điều kiện.*

Các bộ phân loại Naive Bayes khác nhau chủ yếu khác nhau bởi các giả định mà chúng đưa ra liên quan đến phân phối P (xi | y).

**II.3. Bộ phân loại Naive Bayes.**

Bộ phân loại Naive Bayes là tập hợp các thuật toán phân loại dựa trên định lý Bayes. Nó không phải là một thuật toán đơn lẻ mà là một họ các thuật toán mà tất cả chúng đều có chung một nguyên tắc, tức là mọi cặp đặc tính được phân loại là độc lập với nhau.

Bộ phân lớp Naive bayes hay mô hình phân lớp Bayes hoạt động như sau:

* Gọi D là tập dữ liệu huấn luyện, trong đó mỗi phần tử dữ liệu X được biểu diễn bằng một vector chứa n giá trị thuộc tính A1, A2,...,An = {x1,x2,...,xn}
* Giả sử có m lớp C1, C2,..,Cm. Cho một phần tử dữ liệu X, bộ phân lớp sẽ gán nhãn cho X là lớp có xác suất hậu nghiệm lớn nhất. Cụ thể, bộ phân lớp Bayes sẽ dự đoán X thuộc vào lớp Ci nếu và chỉ nếu:  
  P(Ci|X) > P(Cj|X) (1<= i, j <=m, i != j) *Giá trị này sẽ tính dựa trên định lý Bayes.*
* Để tìm xác suất lớn nhất, ta nhận thấy các giá trị P(X) là giống nhau với mọi lớp nên không cần tính. Do đó ta chỉ cần tìm giá trị lớn nhất của P(X|Ci) \* P(Ci). Chú ý rằng P(Ci) được ước lượng bằng |Di|/|D|, trong đó Di là tập các phần tử dữ liệu thuộc lớp Ci. Nếu xác suất tiền nghiệm P(Ci) cũng không xác định được thì ta coi chúng bằng nhau P(C1) = P(C2) = ... = P(Cm), khi đó ta chỉ cần tìm giá trị P(X|Ci) lớn nhất.
* Khi số lượng các thuộc tính mô tả dữ liệu là lớn thì chi phí tính toàn P(X|Ci) là rất lớn, dó đó có thể giảm độ phức tạp của thuật toán Naive Bayes giả thiết các thuộc tính độc lập nhau. Khi đó ta có thể tính:  
  P(X|Ci) = P(x1|Ci)...P(xn|Ci)

**II.4. Các phân phối thường dung cho P(xi|c).**

* *Multinomial Naive Bayes:* Mô hình này chủ yếu được sử dụng trong phân loại văn bản. Đặc trưng đầu vào ở đây chính là tần suất xuất hiện của từ trong văn bản đó.
* *Bernoulli Naive Bayes:* Mô hình này được sử dụng khi các đặc trưng đầu vào chỉ nhận giá trị nhị phân 0 hoặc 1 (phân bố Bernoulli).
* *Gaussian Naive Bayes:* Khi các đặc trưng nhận giá trị liên tục, ta giả sử các đặc trưng đó có phân phối Gaussian.

**PHẦN III: TÌM HIỂU VỀ COVID 19.**

**III.1. Tìm hiểu chung về covid 19.**

COVID-19 là một [bệnh đường hô hấp](https://vi.wikipedia.org/wiki/B%E1%BB%87nh_%C4%91%C6%B0%E1%BB%9Dng_h%C3%B4_h%E1%BA%A5p) cấp tính truyền nhiễm gây ra bởi chủng virus corona [SARS-CoV-2](https://vi.wikipedia.org/wiki/SARS-CoV-2) và [các biến thể của nó](https://vi.wikipedia.org/wiki/C%C3%A1c_bi%E1%BA%BFn_ch%E1%BB%A7ng_c%E1%BB%A7a_SARS-CoV-2). Đây là một loại virus mới phát hiện điều tra ổ dịch bắt nguồn từ khu chợ lớn chuyên bán hải sản và động vật ở [Vũ Hán](https://vi.wikipedia.org/wiki/V%C5%A9_H%C3%A1n), tỉnh [Hồ Bắc](https://vi.wikipedia.org/wiki/H%E1%BB%93_B%E1%BA%AFc), [Trung Quốc](https://vi.wikipedia.org/wiki/Trung_Qu%E1%BB%91c). Virus gây viêm đường hô hấp cấp ở người và cho thấy có sự lây lan từ người sang người. Ngoài chủng virus corona mới phát hiện này, đã có 6 chủng virus corona khác được biết tới ngày nay có khả năng lây nhiễm ở [người](https://vi.wikipedia.org/wiki/Ng%C6%B0%E1%BB%9Di) sang người. Bệnh được phát hiện lần đầu tiên trong [đại dịch COVID-19](https://vi.wikipedia.org/wiki/%C4%90%E1%BA%A1i_d%E1%BB%8Bch_COVID-19) năm 2019–2020.

Phương thức lây truyền chủ yếu của nó hiện nay là lây truyền từ người sang người, thường được truyền thông qua các giọt dịch [hô hấp](https://vi.wikipedia.org/wiki/H%C3%B4_h%E1%BA%A5p_(sinh_l%C3%BD_h%E1%BB%8Dc)) mà con người [hắt hơi](https://vi.wikipedia.org/wiki/H%E1%BA%AFt_h%C6%A1i), [ho](https://vi.wikipedia.org/wiki/Ho) hoặc [thở ra](https://vi.wikipedia.org/wiki/Th%E1%BB%9F_ra). Một người nhiễm bệnh có thể xuất hiện các triệu chứng sau [thời gian ủ bệnh](https://vi.wikipedia.org/wiki/Th%E1%BB%9Di_gian_%E1%BB%A7_b%E1%BB%87nh) kéo dài từ 2 đến 14 ngày, trung bình là 5 ngày, trong thời gian đó nó vẫn có thể truyền nhiễm. Cần thận trọng để giúp hạn chế lây truyền bệnh, bao gồm vệ sinh cá nhân tốt và rửa tay thường xuyên. Những người nghĩ rằng họ đã bị nhiễm bệnh nên đeo [khẩu trang y tế](https://vi.wikipedia.org/wiki/Kh%E1%BA%A9u_trang_y_t%E1%BA%BF) và liên hệ [bác sĩ](https://vi.wikipedia.org/wiki/B%C3%A1c_s%C4%A9) ngay lập tức để được tư vấn.

[Virus corona](https://vi.wikipedia.org/wiki/SARS-CoV-2) chủ yếu ảnh hưởng đến [đường hô hấp dưới](https://vi.wikipedia.org/wiki/%C4%90%C6%B0%E1%BB%9Dng_h%C3%B4_h%E1%BA%A5p) (cũng có các triệu chứng ở đường hô hấp trên nhưng ít gặp hơn) và dẫn đến một loạt các triệu chứng được mô tả giống như [cúm](https://vi.wikipedia.org/wiki/C%C3%BAm), bao gồm [sốt](https://vi.wikipedia.org/wiki/S%E1%BB%91t), [ho](https://vi.wikipedia.org/wiki/Ho), [khó thở](https://vi.wikipedia.org/wiki/Kh%C3%B3_th%E1%BB%9F), [đau cơ](https://vi.wikipedia.org/wiki/%C4%90au_c%C6%A1) và [mệt mỏi](https://vi.wikipedia.org/wiki/M%E1%BB%87t_m%E1%BB%8Fi), với sự phát triển cao hơn nữa sẽ dẫn đến [viêm phổi](https://vi.wikipedia.org/wiki/Vi%C3%AAm_ph%E1%BB%95i), hội chứng suy hô hấp cấp tính, [nhiễm trùng huyết](https://vi.wikipedia.org/wiki/Nhi%E1%BB%85m_tr%C3%B9ng_huy%E1%BA%BFt), sốc nhiễm trùng và có thể gây [tử vong](https://vi.wikipedia.org/wiki/Ch%E1%BA%BFt). Các phản ứng y tế đối với căn bệnh này thường là cố gắng kiểm soát các triệu chứng lâm sàng vì hiện tại chưa tìm thấy phương pháp điều trị hiệu quả nào.

**III.2. Cách thức lây lan và triệu chứng thường gặp.**

***Cách thức lây lan***

Đến nay, bệnh được xác định là có lây truyền từ người sang người qua đường tiếp xúc trực tiếp với chất tiết từ [đường hô hấp](https://vi.wikipedia.org/wiki/%C4%90%C6%B0%E1%BB%9Dng_h%C3%B4_h%E1%BA%A5p) của người bệnh. Bệnh còn lây lan đường gián tiếp khi bàn tay người lành tiếp xúc với các đồ vật bị nhiễm virus như [bao cao su](https://vi.wikipedia.org/wiki/Bao_cao_su), [băng vệ sinh](https://vi.wikipedia.org/wiki/B%C4%83ng_v%E1%BB%87_sinh)... đặc biệt là các đồ vật có dính dịch mũi họng của người bị COVID-19, sau đó đưa tay vào [mắt](https://vi.wikipedia.org/wiki/M%E1%BA%AFt), [mũi](https://vi.wikipedia.org/wiki/M%C5%A9i), [miệng](https://vi.wikipedia.org/wiki/Mi%E1%BB%87ng) và gây nhiễm bệnh. Đó, phòng bệnh cơ bản là rửa tay thường xuyên bằng [xà phòng](https://vi.wikipedia.org/wiki/X%C3%A0_ph%C3%B2ng) và các biện pháp dự phòng không dùng thuốc. Theo một số nghiên cứu, bệnh cũng có thể lây qua [đường tiêu hóa](https://vi.wikipedia.org/wiki/H%E1%BB%87_ti%C3%AAu_h%C3%B3a_ng%C6%B0%E1%BB%9Di). COVID-19 vẫn có khả năng lây nhiễm cao dù người bị bệnh đang trong thời gian ủ bệnh.

***Nguyên nhân gây bệnh***

[Nguyên nhân](https://vi.wikipedia.org/wiki/Nguy%C3%AAn_nh%C3%A2n_(y_h%E1%BB%8Dc)) gây bệnh đã được khẳng định là do chủng virut tên là [SARS-CoV-2](https://vi.wikipedia.org/wiki/SARS-CoV-2) gây ra và được cho là có nguồn gốc [từ động vật](https://vi.wikipedia.org/wiki/%C4%90%E1%BB%99ng_v%E1%BA%ADt_l%C3%A2y_truy%E1%BB%81n_b%E1%BB%87nh). Về mặt Sinh học, khi ở ngoài cơ thể vật chủ, thì CoV-19 không phải là sinh vật, như tất cả các virut khác, do nó chỉ có biểu hiện của vật sống khi [ký sinh](https://vi.wikipedia.org/wiki/K%C3%BD_sinh) trong [vật chủ](https://vi.wikipedia.org/wiki/V%E1%BA%ADt_ch%E1%BB%A7) như người.

**Phương thức lây truyền**

Phương thức lây truyền chính là từ người sang người, qua các giọt dịch hô hấp của cơ thể khi mà người có mầm bệnh [hắt hơi](https://vi.wikipedia.org/wiki/H%E1%BA%AFt_h%C6%A1i), [ho](https://vi.wikipedia.org/wiki/Ho) hoặc [thở ra](https://vi.wikipedia.org/wiki/Th%E1%BB%9F_ra). Các nhà khoa học và quan chức có trách nhiệm đã xác nhận ba cách lây truyền: truyền trực tiếp, truyền tiếp xúc và truyền khí dung.

* Trong cách truyền trực tiếp: các giọt dịch hô hấp (nói nôm na là giọt [nước bọt](https://vi.wikipedia.org/wiki/N%C6%B0%E1%BB%9Bc_b%E1%BB%8Dt)) của người có mầm bệnh bắn trực diện vào cơ quan hô hấp của người khác (thường là [mũi](https://vi.wikipedia.org/wiki/M%C5%A9i), [miệng](https://vi.wikipedia.org/wiki/Mi%E1%BB%87ng), một số trường hợp là [mắt](https://vi.wikipedia.org/wiki/M%E1%BA%AFt)), khi ở gần nhau. Trường hợp này rất hay gặp khi một số người đối diện nhau mà nói chuyện, cười đùa với nhau.
* Trong cách truyền tiếp xúc: các giọt dịch hô hấp của người có mầm bệnh bám trên vật thể nào đó, rồi người khác chạm phải, tay có dính mầm bệnh lại vô tình quệt lên [mắt](https://vi.wikipedia.org/wiki/M%E1%BA%AFt), ngoáy [mũi](https://vi.wikipedia.org/wiki/M%C5%A9i) hay cầm thức ăn cho vào [miệng](https://vi.wikipedia.org/wiki/Mi%E1%BB%87ng). Chẳng hạn như người có mầm bệnh hắt hơi hoặc nhổ lên một tay ghế, người khác nắm hay chạm phải có thể dính mầm bệnh.
* Trong cách truyền khí dung - còn gọi là truyền qua không khí - thì mầm bệnh bay lơ lửng trong [không khí](https://vi.wikipedia.org/wiki/Kh%C3%B4ng_kh%C3%AD), người khác chẳng may hít phải hoặc tiếp xúc qua niêm mạc. Trường hợp này thường gặp khi nhiều người tập trung ở phòng kín (lúc hội họp, tập trung xem biểu diễn) hoặc ngay cả ở ngoài trời (ít gặp hơn, nhưng không phải là khó xảy ra), do mầm bệnh có thể tồn tại trong không khí nhiều giờ, nhất là khi nhiệt độ không khí mát hay lạnh (<18 °C).

***Dấu hiệu và triệu trứng***

Những người bị nhiễm bệnh có thể [không có triệu chứng](https://vi.wikipedia.org/wiki/Ng%C6%B0%E1%BB%9Di_mang_m%E1%BA%A7m_b%E1%BB%87nh_kh%C3%B4ng_tri%E1%BB%87u_ch%E1%BB%A9ng) hoặc có các [triệu chứng cơ năng](https://vi.wikipedia.org/wiki/Tri%E1%BB%87u_ch%E1%BB%A9ng_c%C6%A1_n%C4%83ng) từ nhẹ đến nặng, như sốt, ho và khó thở. [Tiêu chảy](https://vi.wikipedia.org/wiki/Ti%C3%AAu_ch%E1%BA%A3y) hoặc các triệu chứng ở [đường hô hấp trên](https://vi.wikipedia.org/wiki/%C4%90%C6%B0%E1%BB%9Dng_h%C3%B4_h%E1%BA%A5p) (ví dụ như hắt hơi, [sổ mũi](https://vi.wikipedia.org/wiki/S%E1%BB%95_m%C5%A9i), đau họng) ít gặp hơn. Một số trường hợp ở Trung Quốc ban đầu chỉ xuất hiện với [đau ngực](https://vi.wikipedia.org/wiki/%C4%90au_ng%E1%BB%B1c) và [đánh trống ngực](https://vi.wikipedia.org/wiki/%C4%90%C3%A1nh_tr%E1%BB%91ng_ng%E1%BB%B1c). Vào tháng 3 năm 2020, các báo cáo nổi lên chỉ ra rằng [mất khứu giác](https://vi.wikipedia.org/wiki/M%E1%BA%A5t_kh%E1%BB%A9u_gi%C3%A1c) có thể là triệu chứng phổ biến ở những người mắc bệnh nhẹ, mặc dù không phổ biến như báo cáo ban đầu. Các trường hợp có thể tiến triển thành [viêm phổi](https://vi.wikipedia.org/wiki/Vi%C3%AAm_ph%E1%BB%95i) nặng, [suy rối loạn đa tạng](https://vi.wikipedia.org/wiki/H%E1%BB%99i_ch%E1%BB%A9ng_r%E1%BB%91i_lo%E1%BA%A1n_ch%E1%BB%A9c_n%C4%83ng_%C4%91a_c%C6%A1_quan) và tử vong.

*Tham khảo bảng dưới:*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sốt | Ho khan | Mệt mỏi | Tiết đờm | Mất khứu giác | Khó thở | Đau họng | Nhức đầu |
| 88% | 68% | 38% | 33% | 15-30% | 19% | 35% | 14% |

**III.3. Các biểu hiện của người bị dương tính với Covid 19.**

- Khó thở, thở hụt hơi, hoặc ở trẻ em có dấu hiệu thở bất thường: thở rên, rút lõm lồng ngực, phập phồng cánh mũi, khò khè, thở rít thì hít vào.

- Nhịp thở: Đối với người lớn: nhịp thở ≥ 20 lần/phút

- Đối với trẻ em từ 1 đến dưới 5 tuổi: Nhịp thở: ≥ 40 lần/phút; Trẻ từ 5 – dưới 12 tuổi: nhịp thở: ≥ 30 lần/phút (Lưu ý ở trẻ em: đếm đủ nhịp thở trong 1 phút khi trẻ nằm yên không khóc).

- SpO2 ≤ 96% (trường hợp phát hiện chỉ số SpO2 bất thường cần đo lại lần 2 sau 30 giây đến 1 phút, khi đo yêu cầu giữ yên vị trí đo).

- Mạch nhanh > 120 nhịp/phút hoặc < 50 nhịp/phút.

- Huyết áp thấp: huyết áp tối đa < 90 mmHg, huyết áp tối thiểu < 60 mmHg (nếu có thể đo).

- Đau tức ngực thường xuyên, cảm giác bó thắt ngực, đau tăng khi hít sâu.

- Thay đổi ý thức: lú lẫn, ngủ rũ, lơ mơ, rất mệt/mệt lả, trẻ quấy khóc, li bì khó đánh thức, co giật.

- Tím môi, tím đầu móng tay, móng chân, da xanh, môi nhợt, lạnh đầu ngón tay, ngón chân.

- Không thể uống hoặc bú kém/giảm, ăn kém, nôn (ở trẻ em). Trẻ có biểu hiện hội chứng viêm đa hệ thống: sốt cao, đỏ mắt, môi đỏ, lưỡi dâu tây, ngón tay chân sưng phù nổi hồng ban...

- Mắc thêm bệnh cấp tính: sốt xuất huyết, tay chân miệng...

- Đối với người lớn: > 38,5 độ hoặc đau đầu, đau người nhiều: uống mỗi lần 1 viên thuốc hạ sốt như paracetamol 0,5 g, có thể lặp lại mỗi 4-6 h, ngày không quá 4 viên, uống oresol nếu ăn kém/giảm hoặc có thể dùng uống thay nước.

- Đối với trẻ em: > 38,5 độ, uống thuốc hạ sốt như paracetamol liều 10-15 mg/kg/lần, có thể lặp lại mỗi 4-6 h, ngày không quá 4 lần.

**III.4. Dữ liệu của các bệnh nhân nhiễm Covid 19 (Minh họa dựa trên các triệu chứng).**

**(***Các dữ liệu được minh họa thông qua các biểu hiện của người nhiễm Covid 19 thông qua bảng excel đính kèm file)*

**IV. ỨNG DỤNG MÔ HÌNH PHÂN LỚP NAÏVE BAYES TRONG CHUẨN ĐOÁN NGƯỜI MẮC COVID 19 THÔNG QUA CÁC TRIỆU CHỨNG VÀ BIỂU HIỆN.**

**IV.1. Dữ liệu của các bệnh nhân nhiễm Covid 19.**

Dữ liệu thu thập được là của 96 người bao gồm cả những người mắc và người không mắc. Dữ liệu thu thập này bao gồm các thuộc tính như sau:

* Độ tuổi
* Nhịp thở/Phút
* Mạch/Phút
* Nhiệt độ cơ thể (độ C)
* Chỉ số SpO2 (%)

Với mỗi người được kiểm tra sức khỏe, dữ liệu bao gồm tập hợp các chỉ số kể trên và tình trạng bị bênh là dương tính hay không ***(1 là dương tính, 0 là âm tính).*** Về bản chất đây là một bài toán phân loại 2 lớp. Chúng ta có thể hình dung tập dữ liệu này thông qua biểu diễn dưới dạng file CSV được chuyển từ Excel sau đó chuyển thành file txt, trong đó cột cuối cùng chính là tình trạng nhiễm Covid hay không. Các cột từ 2 đến 6 tương ứng với các chỉ số được nêu ở bên trên.

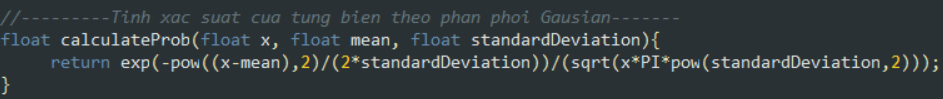
Có một điều nhận thấy rằng giá trị của các chỉ số là một ***biến liên tục*** chứ không phải một giá trị ***rời rạc*** chính vì thế nên khi áp dụng thuật toán ***Naive Bayes*** chúng ta cần phải áp dụng một ***phân phối xác suất*** cho nó. Một trong những phân phối xác suất phổ biến được sử dụng trong phần này đó chính là ***phân phối Gaussian***.

Diagram

Description automatically generated with low confidenceNhư đã giới thiệu về phân phối Gaussian ở phần I, chúng ta sẽ tìm hiểu rõ hơn về nó. Với một một dữ liệu xi​ thuộc một class ci​ chúng ta thấy xi​ tuân theo một phân phối chuẩn với kì vọng μ và độ lệch chuẩn σ. Khi đó hàm xác suất của xi​ được xác định như sau:

**IV.2. Cài đặt thủ công.**

* Text

  Description automatically generatedXây dựng hàm tính giá trị trung bình và độ lệch chuẩn cho những trường hợp âm tính và trường hợp dương tính (Bên dưới là dành cho trường hợp dương tính).
* Sau khi tính được trung bình và độ lệch chuẩn ta tính xác xuất theo phân phối Gaussian theo công thức đã nêu ở trên.****
* Sau đó thức hiện quá trình đọc dữ liệu từ file .txt đã lưu để có các thông số tính toán.
* Lưu các thông số vào các biến phù hợp (Đã thể hiện ở trong code).
* Thực hiện quá trình training tính toán giá trị trung bình, độ lệch chuẩn để tính xác xuất theo phân phối Gaussian đối với mỗi biến.

Text

Description automatically generated

* Tính tích tất các các giá trị lưu vào biến at và dt để so sánh. Biến at và dt được tính bằng cách tính tích xác suất các biến theo phân phối Gaussian với xác xuất dương tính hay không (đã được tính ở các bước trên).
* Nhâp một bộ các thuộc tính mới để test.

**IV.3. Các vấn đề gặp phải.**

* Không thế nhận dạng được toàn bộ trường hợp.
* Kết quá không đúng khi dữ liệu test nhập sai so với thực tế.
* Tỉ lệ đúng chưa cao.
* Giao diện chương trình đơn sơ.

**V.TỔNG KẾT**

*Ưu điểm*: Hoạt động tốt cho nhiều bài toán/miền sữ liệu và ứng dụng.Đơn giản nhưng đủ tốt để giải quyết nhiều bài toán như phân lớp văn bản, lọc spam,… Cho phép kết hợp tri thức tiền nghiệm (prior knowledge) và dữ liệu quan sát được (obserwed data). Tốt khi có sự chệnh lệch số lượng giữa các lớp phân loại.Huấn luyện mô hình (ước lượng tham số) dễ và nhanh.

*Nhược điểm*: Hầu hết các trường hợp thực tế trong đó có các thuộc tính trong các đối tượng thường phụ thuộc lẫn nhau. Vấn đề zero (đã nêu cách giải quyết ở phía trên). Mô hình không được huẩn luyện bằng phượng pháp tối ưu mạnh và chặt chẽ. Tham số mủa mô hình là các ước lượng xác suất điều kiện đơn lẻ.Không tính đến sự tương tác giữa các ước lượng này.