|  |
| --- |
| Mẫu 1  TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT VINH (Cỡ chữ 18)  KHOA CÔNG NGHỆ THÔNG TIN  NGÀNH/CHUYÊN NGÀNH: CÔNG NGHỆ THÔNG TIN  **ĐỒ ÁN TỐT NGHIỆP**  **ỨNG DỤNG KỸ THUẬT HỌC MÁY TRONG**  **CHẨN ĐOÁN BỆNH TIỂU ĐƯỜNG**  ***Giảng viên hướng dẫn*: ThS. Lê Thanh Tươi**  ***Sinh viên thực hiện* : Cao Chí Nhân**  ***Lớp* : ĐHCTTCK14A1**  Nghệ An - 2023 |

**Lời mở đầu**

Mẫu 2

Đái tháo đường là một trong những bệnh không lây nhiễm phổ biến trên toàn cầu. Năm 2019, trên toàn thế giới có 463 triệu người lớn (độ tuổi 20-79) tương đương 1 trong 11 người trưởng thành đang sống với bệnh đái tháo đường. Dự đoán vào năm 2045, con số này sẽ tăng tới khoảng 700 triệu người, hay nói cách khác 1 người trong 10 người lớn sẽ có bệnh đái tháo đường. Tuy nhiên, gần một nửa số người đang sống với bệnh đái tháo đường (độ tuổi 20-79) không được chẩn đoán (46,5%), tỷ lệ này ở khu vực Tây Thái Bình Dương là 52.1%. Ước tính hơn 4 triệu người trong độ tuổi từ 20-79 đã tử vong vì các nguyên nhân liên quan đến đái tháo đường trong năm 2019. Bệnh đái tháo đường gây nên nhiều biến chứng nguy hiểm, là nguyên nhân hàng đầu gây bệnh tim mạch, mù lòa, suy thận, và cắt cụt chi.[1](https://thuvienphapluat.vn/van-ban/The-thao-Y-te/Quyet-dinh-5481-QD-BYT-2020-tai-lieu-chuyen-mon-Huong-dan-chan-doan-dieu-tri-dai-thao-duong-tip-2-460925.aspx" \l "_ftn1" \o ")

Để tăng cường chất lượng và chuẩn hóa công tác chuyên môn chẩn đoán, điều trị đái tháo đường, năm 2017, Bộ Y tế đã ban hành Hướng dẫn chẩn đoán và điều trị đái tháo đường. Với tiến bộ khoa học kỹ thuật, khuyến cáo cập nhật năm 2020 của Hiệp Hội Đái tháo đường thế giới, Hiệp hội Đái tháo đường Hoa kỳ về hướng dẫn chăm sóc, điều trị đái tháo đường, Bộ Y tế đã triển khai cập nhật Hướng dẫn chẩn đoán và điều trị đái tháo đường. Hướng dẫn được xây dựng công phu, cập nhật, dựa trên các tài liệu trong nước và quốc tế, tập trung chủ yếu vào thực hành lâm sàng chẩn đoán và điều trị đái tháo đường do vậy sẽ rất hữu ích cho các thầy thuốc đa khoa, chuyên khoa trong hành nghề khám, chữa bệnh hàng ngày. Đây là bản được cập nhật lần thứ tư của Bộ Y tế về đái tháo đường.

Để giảm thiểu nguy cơ mắc bệnh tiểu đường hiện nay chúng ta có thể đưa phương pháp học máy machine learning vào để chẩn đoán bệnh, giúp cho việc điều trị và chữa trị được sớm hơn, giảm thiểu tối đa các biến chứng mà bệnh gây ra.

**Tóm tắt nội dung**

- Vấn đề cần thực hiện: Thu thập các thông tin, dữ liệu về người bị bệnh tiểu đường qua bệnh viện, internet,… về các thông số, chỉ số để chuẩn đoán con người có mắc bệnh tiểu đường hay không thông qua phương pháp học machine learning (học máy).

- Phương pháp thực hiện:

+ Thu thập dữ liệu, thông tin người bệnh qua bệnh viện, internet,…

+ Sử dụng phần mềm: Weka, Python, Pycharm.

+ Kết quả đạt được: Dự đoán chuẩn 80% với vấn đề dự đoán bệnh tiểu đường so với dữ liệu người bệnh mà ta đem vào.

+ Tính thực tế: Đồ án này có tính thực tế rất cao với nhu cầu về sức khỏe của con người hiện đại ngày nay.

+ Các kiến thức và kỹ năng mà sinh viên đạt được: Hiểu biết thêm về việc thu thập dữ liệu thực tế, kỹ năng lập trình, hiểu biết thêm về các phần mềm để ứng dụng của phương pháp học máy machine learning như: weka, python, pycharm, excel,…

*Nghệ An, ngày 09 tháng 3 năm 2023*

**Sinh viên thực hiện**

*(Ký và ghi rõ họ tên)*

Cao Chí Nhân

|  |
| --- |
| **Nhận xét của giảng viên hướng dẫn**  Mẫu 3  ………………………………………………………………………………………  ………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………  *Nghệ An, ngày 09 tháng 3 năm 2023*  **Giảng viên hướng dẫn**  *(Ký và ghi rõ họ tên)* |

**CHƯƠNG 1: TỔNG QUAN VỀ MACHINE LEARING**

**1.1 Định nghĩa:**

Machine learning là một nhánh của trí tuệ nhân tạo (AI) và khoa học máy tính, tập trung vào việc sử dụng dữ liệu và thuật toán để bắt chước hành động của con người, dần dần cải thiện độ chính xác của nó. Machine learning còn là một thành phần quan trọng của lĩnh vực khoa học dữ liệu đang phát triển

[**1.2 Các**](#_Toc21499668) **phương pháp học máy machine learning:**

[1.2.1 Học có giám sát.](#_Toc21499669)

### [\* Supervised Learning (Học có giám sát)](#_Toc21499669)

[Supervised learning là thuật toán dự đoán đầu ra (outcome) của một dữ liệu mới (new input) dựa trên các cặp (input, outcome) đã biết từ trước. Cặp dữ liệu này còn được gọi là (data, label), tức (dữ liệu, nhãn). Supervised learning là nhóm phổ biến nhất trong các thuật toán Machine Learning.](#_Toc21499669)

[Một cách toán học, Supervised learning là khi chúng ra có một tập hợp biến đầu vào X={x1,x2,…,xN} và một tập hợp nhãn tương ứng Y={y1,y2,…,yN}, trong đó xi,yi là các vector. Các cặp dữ liệu biết trước (xi,yi)∈X×Y được gọi là tập training data (dữ liệu huấn luyện). Từ tập training data này, chúng ta cần tạo ra một hàm số ánh xạ mỗi phần tử từ tập X sang một phần tử (xấp xỉ) tương ứng của tập Y:](#_Toc21499669)

[yi≈f(xi),  ∀i=1,2,…,N,](#_Toc21499669)

[Mục đích là xấp xỉ hàm số f thật tốt để khi có một dữ liệu x mới, chúng ta có thể tính được nhãn tương ứng của nó y=f(x).](#_Toc21499669)

**[Ví dụ :](#_Toc21499669)**[Bản thân thuật toán dò tìm các khuôn mặt trong 1 bức ảnh cũng là một thuật toán Supervised learning với training data (dữ liệu học) là hàng ngàn cặp (ảnh, mặt người) và (ảnh, không phải mặt người) được đưa vào. Chú ý là dữ liệu này chỉ phân biệt mặt người và không phải mặt người mà không phân biệt khuôn mặt của những người khác nhau.](#_Toc21499669)

[1.2.2 Học không giám sát.](#_Toc21499671)

### [\* Unsupervised Learning (Học không giám sát)](#_Toc21499671)

[Trong thuật toán này, chúng ta không biết được outcome hay nhãn mà chỉ có dữ liệu đầu vào. Thuật toán unsupervised learning sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó, ví dụ như phân nhóm (clustering) hoặc giảm số chiều của dữ liệu (dimension reduction) để thuận tiện trong việc lưu trữ và tính toán.](#_Toc21499671)

[Một cách toán học, Unsupervised learning là khi chúng ta chỉ có dữ liệu vào X mà không biết nhãn Y tương ứng.](#_Toc21499671)

[Những thuật toán loại này được gọi là Unsupervised learning vì không giống như Supervised learning, chúng ta không biết câu trả lời chính xác cho mỗi dữ liệu đầu vào. Giống như khi ta học, không có thầy cô giáo nào chỉ cho ta biết đó là chữ A hay chữ B. Cụm không giám sát được đặt tên theo nghĩa này.](#_Toc21499671)

[1.2.3 Học bán giám sát](#_Toc21499672)

[\* Semi-Supervised Learning (Học bán giám sát)](#_Toc21499672)

[Các bài toán khi chúng ta có một lượng lớn dữ liệu X nhưng chỉ một phần trong chúng được gán nhãn được gọi là Semi-Supervised Learning. Những bài toán thuộc nhóm này nằm giữa hai nhóm được nêu bên trên.](#_Toc21499672)

[Một ví dụ điển hình của nhóm này là chỉ có một phần ảnh hoặc văn bản được gán nhãn (ví dụ bức ảnh về người, động vật hoặc các văn bản khoa học, chính trị) và phần lớn các bức ảnh/văn bản khác chưa được gán nhãn được thu thập từ internet. Thực tế cho thấy rất nhiều các bài toán Machine Learning thuộc vào nhóm này vì việc thu thập dữ liệu có nhãn tốn rất nhiều thời gian và có chi phí cao. Rất nhiều loại dữ liệu thậm chí cần phải có chuyên gia mới gán nhãn được (ảnh y học chẳng hạn). Ngược lại, dữ liệu chưa có nhãn có thể được thu thập với chi phí thấp từ internet.](#_Toc21499672)

1.2.4 Học củng cố

### \* Reinforcement Learning (Học Củng Cố)

Reinforcement learning là các bài toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất (maximizing the performance). Hiện tại, Reinforcement learning chủ yếu được áp dụng vào Lý Thuyết Trò Chơi (Game Theory), các thuật toán cần xác định nước đi tiếp theo để đạt được điểm số cao nhất.

**[1.3 Ứng dụng Machine Learning trong lĩnh vực y tế](#_Toc21499677)**

*[\* Xác định và chuẩn đoán bệnh:](#_Toc21499677)*

[Một trong những ứng dụng ML chính trong chăm sóc sức khỏe là xác định và chẩn đoán bệnh, nhất là các bệnh được cho là khó chẩn đoán, từ ung thư giai đoạn sớm đến các bệnh di truyền khác. “IBM Watson Genomics” là một ví dụ điển hình về cách tích hợp điện toán nhận thức với giải trình tự khối u dựa trên bộ gen có thể giúp chẩn đoán nhanh. Berg (công ty chuyên nghiên cứu về sinh học và thuốc) đang tận dụng trí tuệ nhân tạo (AI) để pát triển các phương pháp trị liệu trong các lĩnh vực như ung thư “PReDicT” (Predicting Response to Depression Treatment) của P1vital phát triển ứng dụng để chẩn đoán và cung cấp điều trị trong các điều kiện lâm sàng thông thường của bệnh trầm cảm.](#_Toc21499677)

*[\* Khám phá và sản xuất thuốc.](#_Toc21499677)*

[Một trong nhưng ứng dụng lâm sàng đầu tiên của máy học nằm trong quy trình khám phá thuốc ở giai đoạn đầu. Bao gồm các công nghệ R&D như giải trình tự thế hệ tiếp theo và y học cá thể có thể giúp tìm ra các điều trị thay thế cho các bệnh lý có nhiều yếu tố. Hiện tại, các kỹ thuật máy học theo kiểu học không giám sát (unsupervised learning) có thể xác định các mẫu trong dữ liệu mà không cung cấp bất kỳ dự đoán nào. Dự án “Hanover” Microsoft phát triển đang sử dụng các công nghệ dựa trên ML cho ra nhiều sáng kiến bao gồm phát triển công nghệ dự trên AI để điều trị ung thư và cá thể hóa kết hợp thuốc cho bệnh ung thư máu dòng tủy cấp AML (Acute Myeloid Leukemia).](#_Toc21499677)

*[\* Chẩn đoán hình ảnh.](#_Toc21499677)*

[Máy học và học sâu đều chịu trách nhiệm cho công nghệ đột phá mang tên “Computer Vision”. Sáng kiến “InnerEye” do Microsoft phát triển, có thể hoạt động trên các công cụ chẩn đoán hình ảnh để phân tích hình ảnh. Khi máy học trở nên dễ tiếp cận hơn và nhất là khi phát triển khả năng giải thích, hy vọng sẽ tìm thấy nhiều nguồn dữ liệu hơn từ các hình ảnh khác nhau.](#_Toc21499677)

*[\* Y học cá thể (Pesonalized Medicine)](#_Toc21499677)*

[Các phương pháp điều trị cá thể không những hiệu quả hơn nhờ sự kết hợp sức khỏe cá thể với các phân tích dự đoán mà còn để nghiên cứu thêm và đánh giá bệnh tốt hơn. Hiện tại, các bác sĩ bị giới hạn trong việc lựa chon từ bộ chẩn đoán cụ thể hoặc ước tính rủi ro cho bệnh nhân dựa trên bệnh sử và thông tin di truyền có sẵn. Nhưng máy học đang có những bước tiến lớn và “IBM Watson Oncology” đã đi đầu trong lĩnh vực này bằng cách tận dụng tiền sử của bệnh nhân để tạo ra nhiều lựa chọn điều trị. Trong thời gian tới, sẽ có nhiều thiết bị và cảm biến sinh học với khả năng đo lường sức khỏe tinh vi được tung ra thị trường, cho phép có nhiều dữ liệu hơn cho các công nghệ chăm sóc sức khỏe.](#_Toc21499677)

*[\* Sửa đổi hành vi dựa trên máy học.](#_Toc21499677)*

[Sửa đổi hành vi là một phần quan trong của y tế dự phòng, từ khi ứng dụng máy học trở nên phổ biến trong chăm sóc sức khỏe, đã có nhiều công ty khởi nghiệp đang tiến hành các ứng dụng của máy học trong lĩnh vực phòng ngừa và nhận dạng ung thư, điều trị,…Somatix là một công ty phân tích dữ liệu dựa trên B2B2C, đã triển khai một ứng dụng dựa trên ML để nhận ra các cử chỉ mà chúng ta thực hiện trong cuộc sống hàng ngày, từ đó cho phép con người biết được những hành vi vô thức của mình và thực hiện các thay đổi cần thiết.](#_Toc21499677)

*[\* Hồ sơ sức khỏe thông minh](#_Toc21499677)*

[Duy trì hồ sơ sức khỏe luôn được cập nhật thật sự là một thách thức, phần lớn các quy trình cập nhật mất rất nhiều thời gian để hoàn thành. Vai trò chính của máy học trong chăm sóc sức khỏe là giảm bớt các quy trình để tiết kiệm thời gian, công sức và tiền bạc. Các phương pháp phân loại tài liệu bằng “vector machines” và kỹ thuật nhận dạng OCR dựa trên máy học như “API Cloud Vision” của Google và công nghệ nhận dạng chữ viết tay dựa trên máy học của MATLAB.MIT ngày nay đang đi đầu trong việc phát triển thế hệ tiếp theo của hồ sơ sức khỏe thông minh giúp chẩn đoán đề xuất điều trị lâm sàng.](#_Toc21499677)

*[\* Thử nghiệm và nghiên cứu lâm sàng.](#_Toc21499677)*

[Máy học có một số ứng dụng tiềm năng trong lĩnh vực thử nghiệm và nghiên cứu lâm sàng. Nhất là trong lĩnh vực dược, các thử nghiệm lâm sàng tốn rất nhiều thời gian và tiền bạc và nhiều trường hợp có thể mất nhiều năm để hoàn thành. Áp dụng phân tích dự đoán dựa trên ML để xác định các ứng viên thử nghiệm lâm sàng tiềm năng có thể giúp các nhà nghiên cứu rút ra một nhóm từ nhiềm điểm dữ liệu khác nhau, chẳng hạn như những lần đi khám bệnh trước đây, phương tiện truyền thông xã hội,… truy cập dữ liệu của những người tham gia thử nghiệm, tìm cờ mẫu tốt nhất để kiểm tra và tận dụng sức mạnh của hồ sơ điện tử để giảm lỗi dựa trên dữ liệu.](#_Toc21499677)

*[\* Thu thập dữ liệu đám đông.](#_Toc21499677)*

[Dữ liệu đám đông là tất cả các dữ liệu trong lĩnh vực y tế hiện nay, cho phép các nhà nghiên cứu và thực hành truy cập vào một lượng lớn thông tin được tải lên bởi mọi người và được sự đồng ý của chính họ. “ResearchKit” của Apple cho phép người dùng truy cập các ứng dụng tương tác áp dụng nhận dạng khuôn mặt dựa trên ML để thử và điều trị bệnh Asperger và Parkinson. IBM gần đây đã hợp tác với Medtronic để giải mã, tích lũy và cung cấp dữ liệu về bệnh tiểu đường và insulin theo thời gian thực dựa trên thông tin được cộng đồng cung cấp. Với những tiến bộ được thực hiện trong IoT, lĩnh vực chăm sóc sức khỏe vẫn đang khám phá những cách mới để sử dụng dữ liệu này và giải quyết các trường hợp khó chẩn đoán và giúp cải thiện tổng thể và chẩn đoán và thuốc.](#_Toc21499677)

*[\* Xạ trị tốt hơn.](#_Toc21499677)*

[Một trong những ứng dụng máy học được nghiên cứu nhiều nhất là lĩnh vực X quang. Phân tích hình ảnh y tế có nhiều biến số riêng biệt có thể phát sinh tại bất kỳ thời điểm cụ thể nào. Có rất nhiều tổn thương, các ổ ung thư,… không thể được mô hình đơn giản hóa bằng các phương trình phức tạp. Vì các thuật toán dựa trên ML học được từ vô số các mẫu khác nhau có sẵn, nên việc chẩn đoán và tìm các biến trở nên dễ dàng hơn. Một trong những ứng dụng phổ biến nhất của máy học trong phân tích hình ảnh là phân loại các đối tượng như tổn thương thành các loại bình thường hoặc bất thường, tổn thương hoặc không tổn thương,…””Google Deepmind Health](#_Toc21499677)

[Đang tích cực giúp các nhà nghiên cứu của UCLH phát triển các thuật toán có thể phát hiện sự khác biệt giữa mô khỏe mạnh và mô ung thư và cải thiện hiệu quả của điều trị bức xạ.](#_Toc21499677)

*[\* Dự đoán bùng phát dịch bệnh.](#_Toc21499677)*

[Các công nghệ dựa trên AI và máy học ngày nay cũng được đưa vào sử dụng trong việc theo dõi và dự đoán dịch bệnh trên toàn thế giới. Các nhà khoa học truy cập vào một lượng lớn dữ liệu được thu thập từ vệ tinh, cập nhật phương tiện truyền thông xã hội thời gian thực, thông tin trang Web,… Mạng lưới thần kinh nhân tạo giúp đối chiếu thông tin này và dự đoán mọi thứ từ dịch sốt rét đến các bệnh truyền nhiễm nghiêm trọng. Dự đoán những đợt bùng phát này đặc biệt hữu ích ở các nước thế giới thứ ba vì thiếu cơ sở hạ tầng y tế. “ProMED-mail” là nền tảng báo cáo dựa trên internet để theo dõi các bệnh đang phát triển và các bệnh mới nổi và cung cấp các báo cáo ổ dịch theo thời gian thực.](#_Toc21499677)

[1.3.1 Các nghiên cứu liên quan:](#_Toc21499677)

* [Phạm Minh Hải, Nguyễn Ngọc Quang,](#_Toc21499677)[[Khái niệm về phương pháp random forest trong cuộc cách mạng machine learning và định hướng ứng dụng trong lĩnh vực viễn thám](#_Toc21499677)](https://jgac.vn/journal/article/view/344)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 39 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/42)
* [Đỗ Thị Nhung, Nguyễn Thị Diễm My, Phạm Văn Mạnh, Phạm Vũ Đông, Bùi Quang Thành, Nghiêm Văn Tuấn, Phạm Minh Hải,](#_Toc21499677)[[Nghiên cứu mô hình phát hiện rác thải nhựa ven biển sử dụng ảnh máy bay không người lái và mạng nơ-ron tích chập sâu](#_Toc21499677)](https://jgac.vn/journal/article/view/543)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 49 (2021)](#_Toc21499677)](https://jgac.vn/journal/issue/view/63)
* [Phạm Văn Mạnh, Nguyễn Ngọc Thạch, Lưu Thị Phương Mai, Bùi Quang Thành, Phạm Minh Tâm, Phạm Minh Hải,](#_Toc21499677)[[Nghiên cứu định lượng giá trị dịch vụ hệ sinh thái trong bối cảnh đô thị hóa tại khu vực thành phố Huế giai đoạn 1995-2018 trên cơ sở dữ liệu viễn thám và GIS](#_Toc21499677)](https://jgac.vn/journal/article/view/349)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 39 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/42)
* [Lưu Thế A nh, Phạm Minh Hải, Vũ Thị Hồng Hà, Kiều Thị Thảo, Nguyễn Ngọc Thắng, Nguyễn Thanh Bình,](#_Toc21499677)[[Đánh giá ảnh hưởng của đô thị hóa đến đất nông nghiệp ở đồng bằng sông Hồng sử dụng tư liệu viễn thám đa thời gian và GIS](#_Toc21499677)](https://jgac.vn/journal/article/view/311)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 40 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/38)
* [Đỗ Thị Ngọc Ánh, Nguyễn Thị Thảo Vân, Nguyễn Phương Anh, Phạm Minh Hải, Hoàng Anh Lê, Bùi Quang Thành, Phạm Văn Mạnh,](#_Toc21499677)[[Giám sát lúa và ước tính sinh khối dựa trên thuật toán học máy với dữ liệu Sentinel-1a đa thời gian](#_Toc21499677)](https://jgac.vn/journal/article/view/540)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 49 (2021)](#_Toc21499677)](https://jgac.vn/journal/issue/view/63)
* [Phạm Văn Mạnh, Phạm Minh Hải, Đỗ Thị Ngọc Ánh, Nguyễn Ngọc Thạch,](#_Toc21499677)[[Ứng dụng phương pháp viễn thám và trắc lượng hình thái trong phân tích ảnh hưởng của thay đổi lớp phủ thực vật và phân mảnh môi trường sống](#_Toc21499677)](https://jgac.vn/journal/article/view/46)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 46 (2020)](#_Toc21499677)](https://jgac.vn/journal/issue/view/9)
* [Phạm Văn Mạnh, Nguyễn Ngọc Thạch, Bùi Quang Thành, Phạm Vũ Đông, Phạm Minh Hải,](#_Toc21499677)[[Tác động của mở rộng đô thị đến di sản văn hóa: nghiên cứu ở khu vực quần thể di tích Huế](#_Toc21499677)](https://jgac.vn/journal/article/view/310)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 40 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/38)
* [Phạm Minh Hải, Vũ Ngọc Phan,](#_Toc21499677)[[Ứng dụng công nghệ viễn thám và GIS nghiên cứu vật liệu cháy trong các kiểu rừng phục vụ công tác phòng chống cháy rừng tại tỉnh Bắc Giang](#_Toc21499677)](https://jgac.vn/journal/article/view/264)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 35 (2018)](#_Toc21499677)](https://jgac.vn/journal/issue/view/33)
* [Nguyễn Hà Phú, Phạm Minh Hải, Nguyễn Trọng Trường Sơn,](#_Toc21499677)[[Đo sâu địa hình đáy biển vùng nước nông bằng phương pháp đo ảnh lập thể sử dụng dữ liệu Worldview-2, thử nghiệm tại bãi Hải Sâm, quần đảo Trường Sa](#_Toc21499677)](https://jgac.vn/journal/article/view/355)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 42 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/43)
* [Phạm Minh Hải, Đỗ Thị Hoài,](#_Toc21499677)[[Nghiên cứu cơ sở khoa học trong lựa chọn chỉ số cảnh quan phù hợp phục vụ công tác giám sát biến đổi cấu trúc rừng ngập mặn. Thực nghiệm tại rừng ngập mặn Mũi Cà Mau](#_Toc21499677)](https://jgac.vn/journal/article/view/353)[,](#_Toc21499677)[[Tạp chí Khoa học Đo đạc và Bản đồ: Số 42 (2019)](#_Toc21499677)](https://jgac.vn/journal/issue/view/43)[.](#_Toc21499677)

1.3.2. Những thách thức.

Các thuật toán được giám sát làm việc với dữ liệu được dán nhãn và có thể được sử dụng cho các hoạt động dự đoán và kê đơn khác nhau (dự đoán giá, đề xuất sản phẩm).

Các thuật toán không giám sát làm việc với dữ liệu chưa được gắn nhãn và được sử dụng để gắn nhãn và phân đoạn nó.

**[CHƯƠNG 2: CHẨN ĐOÁN BỆNH DỰA TRÊN HỌC MÁY](#_Toc21499684)**

2.1. Giới thiệu bài toán

Bài toán đặt ra là dự đoán một người có khả năng bị bệnh tiểu đường không thông qua các thông tin mà người lập trình thu thập được đem vào cho máy học, máy sẽ dự đoán người đó có mắc bệnh không qua các chỉ số người dùng nhập vào thông qua bằng phần mềm weka, sau đó lập trình ra giao diện, ứng dụng đưa vào sử dụng.

[2.1.1 S](#_Toc21499686)ơ lược về bệnh tiểu đường.

## **Bệnh tiểu đường là nhóm bệnh lý nội khoa, do rối loạn chuyển hóa Insulin trong cơ thể, dẫn đến lượng đường trong máu tăng cao. Cần xác định nguyên nhân bệnh tiểu đường chính là bước cần thiết để lựa chọn cách chăm sóc và chữa bệnh hiệu quả nhất.**

Bệnh tiểu đường, theo y học còn gọi là bệnh đái tháo đường, là một bệnh rối loạn chuyển hóa mạn tính rất phổ biến. Khi mắc bệnh tiểu đường, cơ thể bạn mất đi khả năng sử dụng hoặc sản xuất ra hormone insulin một cách thích hợp. Mắc bệnh tiểu đường có nghĩa là bạn có lượng đường trong máu quá cao do nhiều nguyên nhân. Tình trạng này có thể gây ra các vấn đề nghiêm trọng cho cơ thể, bao gồm cả mắt, thận, thần kinh và tim. Cần xác định nguyên nhân bệnh tiểu đường chính là bước cần thiết để lựa chọn cách chăm sóc và chữa bệnh hiệu quả nhất.

[2.1.2](#_Toc21499687) Nguyên nhân dẫn đến bệnh tiểu đường là gì ?

Để biết được nguyên nhân gây ra bệnh tiểu đường, đầu tiên bạn phải hiểu được quá trình chuyển hóa glucose trong cơ thể.

Trao đổi glucose là chất cần thiết cho cơ thể của bạn, nó đóng vai trò là nguồn năng lượng cho các tế bào của cơ bắp và các mô, đặc biệt não bộ. Glucose có trong thực phẩm bạn ăn và được dự trữ trong gan (tạo thành glycogen). Trong trường hợp bạn biếng ăn dẫn đến lượng glucose trong máu quá thấp, gan của bạn sẽ ly giải những phân tử glycogen thành glucose và cân bằng lại lượng đường trong máu. Máu hấp thụ glucose và cung cấp cho các tế bào trong cơ thể. Tuy nhiên, các tế bào này không thể sử dụng nguồn “nhiên liệu” này một cách trực tiếp, mà phải có sự hỗ trợ của hormone insulin (được sản xuất bởi tuyến tụy). Sự có mặt của insulin cho phép glucose được hấp thu vào các tế bào, làm giảm nồng độ glucose trong máu. Sau đó, khi đường huyết của bạn đã giảm, tuyến tụy cũng sẽ giảm sản xuất insulin.

Bất kỳ sự bất thường nào xảy ra trong quá trình trao đổi chất này đều có thể làm cho glucose không thể đi vào các tế bào và cung cấp năng lượng. Kết quả là lượng đường vẫn còn trong máu. Sự mất cân bằng này được tích lũy qua thời gian và dẫn đến lượng đường trong máu cao đáng kể, được gọi là tăng đường huyết.

[2.1.3 P](#_Toc21499688)hân loại bệnh tiểu đường

Bệnh tiểu đường được phân làm 3 loại chính, đó là:

Tiểu đường type 1, Tiểu đường type 2, và tiểu đường thai kỳ.

- Bệnh tiểu đường type 1 là bệnh rối loạn tự miễn, trong đó hệ thống miễn dịch của cơ thể tấn công tuyến tụy chứ không phải các yếu tố bên ngoài. Điều này gây ra việc thiếu hụt Insulin và tăng lượng đường huyết trong máu

Nếu bạn bị tiểu đường type 1, các triệu chứng sẽ xuất hiện từ sớm. Hiện nay, vẫn chưa xác định được nguyên nhân tiểu đường type 1. Tuy nhiên,  các bác sĩ cho rằng nguyên nhân tiểu đường type 1 có thể từ sự di truyền hoặc do môi trường. Bạn có nguy cơ mắc tiểu đường type 1 cao hơn nếu rơi vào những trường hợp sau đây:

- Mẹ hoặc anh chị mắc tiểu đường type 1

- Trong cơ thể hiện diện kháng thể bệnh tiểu đường

- Thiếu vitamin D, sớm sử dụng sữa bò hoặc các sữa bột nguồn gốc từ sữa bò, và sử dụng các loại ngũ cốc trước 4 tháng tuổi

**- Bệnh tiểu đường type 2**

Đây là bệnh tiểu đường không phụ thuộc vào Insulin, là loại tiểu đường phổ biến nhất hiện nay. Bệnh thường xuất hiện ở người trưởng thành, tuy nhiên, do tỷ lệ béo phì ngày càng tăng dẫn đến tỷ lệ béo phì ở trẻ vị thành niên ngày một tăng cao

Khi mắc tiểu đường type 2, các tế bào của bạn kháng thể Insulin, tuyến tụy không thể sản xuất Insulin để cung cấp đủ cho cơ thể. Thay vì di chuyển để tạo ra năng lượng cho cơ thể, đường sẽ tích tụ trong máu.

**- Tiểu đường thai kỳ**

Bệnh tiểu đường thai kỳ là một loại tiểu đường xảy ra trong quá trình mang thai của phụ nữ. Bệnh này có thể gây ra các vấn đề cho cả mẹ và trẻ sơ sinh. Tuy nhiên tiểu đường thai kỳ thường sẽ biến mất sau khi mẹ bầu chuyển dạ.

2.1.4 Các triệu chứng của bệnh tiểu đường là gì ?

Các triệu chứng của bệnh tiểu đường bao gồm:

* Có cảm giác cực kỳ khát, hay còn được gọi là chứng khát nhiều;
* Đi tiểu nhiều, đôi khi đi thường xuyên mỗi giờ, còn gọi là chứng tiểu nhiều;
* Sụt cân không rõ nguyên nhân;
* Cảm thấy kiệt sức và mệt mỏi.

Bạn có thể có hoặc không có các triệu chứng bệnh tiểu đường khác, như:

* Buồn nôn hoặc nôn mửa;
* Mờ mắt;
* Nhiễm trùng âm đạo thường xuyên ở phụ nữ;
* Nhiễm nấm men hoặc nấm candida;
* Khô miệng;
* Chậm lành vết loét hoặc vết cắt;
* Ngứa da, đặc biệt là ở bẹn hoặc khu vực âm đạo.

2.1.5 Khi nào bạn nên đi bác sỹ ?

Bạn nên đến bệnh viện để kiểm tra hoặc gặp bác sĩ nếu bạn nhận thấy bất kỳ triệu chứng nào của bệnh tiểu đường đã đề cập ở trên.

Ngoài ra, bạn cần phải gọi cấp cứu ngay nếu bạn:

* Có cảm giác buồn nôn và yếu tay chân;
* Cảm thấy khát nhiều hoặc đi tiểu thường xuyên kèm với đau vùng bụng;
* Thở gấp hơn

Bệnh tiểu đường là căn bệnh có thể gặp ở mọi đối tượng, đặc biệt hiện nay căn bệnh này còn có nguy cơ trẻ hóa. Nhận thấy được mối nguy hiểm đó, hiện nay đã và đang tiếp tục triển khai dịch vụ khám và sàng lọc bệnh tiểu đường, rối loạn mỡ máu nhằm giúp phát hiện sớm tình trạng tiền đái tháo đường, phân loại chính xác type tiểu đường, đồng thời xây dựng chế độ dinh dưỡng, theo dõi giảm thiểu các nguy cơ, biến chứng do tiểu đường gây ra.

2.2 Phương pháp học máy áp dụng trong bài toán.

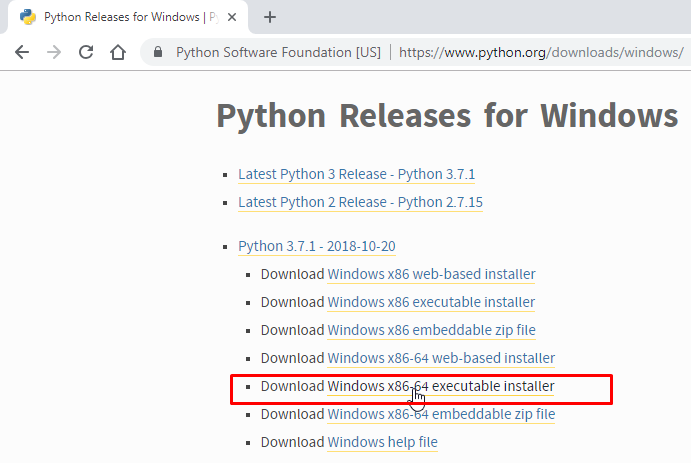
[CHƯƠNG 3.](#_Toc21499691) CÀI ĐẶT THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1 Cài đặt thử nghiệm.

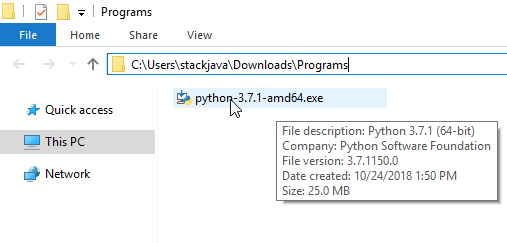
**Download và cài đặt Python**

Download file cài đặt python tại[**https://www.python.org/downloads/windows/**](https://www.python.org/downloads/windows/)

Chọn version mà bạn muốn cài đặt. Ở đây mình cài version 3.7 bản **Windows x84-64 executable installer**

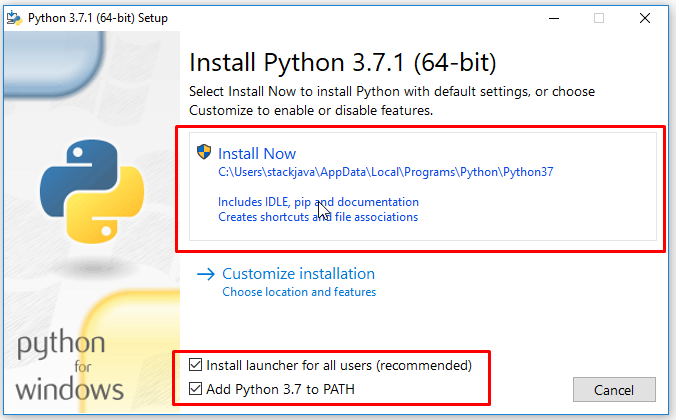
[](https://codecute.com/wp-content/uploads/2018/10/window-install-python-1.png)

Click đúp vào file .exe vừa tải về để cài đặt

[](https://codecute.com/wp-content/uploads/2018/10/window-install-python-2.png)

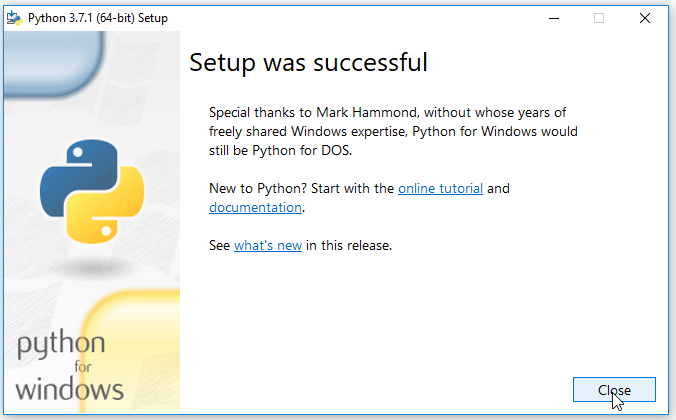
Chọn tick vào ô Add Python 3.7 to PATH (lựa chọn này giúp bạn chạy lệnh python trên cmd, powershell trên windows)

Và click vào Install Now (Bạn có thể đổi đường dẫn folder cài đặt bằng cách chọn phần Customize installation ở dưới)

[](https://codecute.com/wp-content/uploads/2018/10/window-install-python-4.png)

Ngồi chờ python cài đặt.

Kết quả:

[](https://codecute.com/wp-content/uploads/2018/10/window-install-python-5.png)

[3.2 45](#_Toc21499693)

KẾT LUẬN ………………………………………………………………….50

DANH MỤC [TÀI LIỆU THAM KHẢO](#_Toc21499694)

[PHỤ LỤC](#_Toc21499695)

Mẫu 5

**DANH MỤC BẢNG BIỂU, HÌNH VẼ, SƠ ĐỒ**

[Hình 1.1 . 20](#_Toc20580104)

[Hình 1.2 . 21](#_Toc20580105)

[Hình 1.3 . 22](#_Toc20580106)

[Hình 1.4 . 23](#_Toc20580107)

[Hình 1.5 . 23](#_Toc20580108)

[Hình 2.1 . 24](#_Toc20580104)

[Hình 2.2 . 26](#_Toc20580105)

[Hình 2.3 . 27](#_Toc20580106)

[Hình 2.4 . 28](#_Toc20580107)

[Hình 2.5 .](#_Toc20580108).............................................................................................................. 29

[Hình 3.1 . 30](#_Toc20580104)

[Hình 3.2 . 31](#_Toc20580105)

[Hình 3.3 . 32](#_Toc20580106)

[Hình 3.4 . 22](#_Toc20580107)

[Hình 3.5 . 34](#_Toc20580108)

**DANH MỤC TÀI LIỆU THAM KHẢO**

Mẫu 6

**Tiếng Việt**

[1]. Nguyễn Đức Cường (2012), *Tổng quan về khai phá dữ liệu*, Khoa CNTT – Đại học Bách Khoa TP HCM.

[2]. Hà Quang Thụy (2010), *Giáo trình kho dữ liệu và khai phá dữ liệu*, Đại học Quốc gia Hà nội.

[3]. Nguyễn Đức Thuần (2013), *Nhập môn khai phá dữ liệu và quản trị tri thức*, NXB Thông tin và truyền thông.

[4]. Rokach Lior, Maimon O., *Data mining with decision tree: theory and applications*, Word Scientific Pub Co Inc. ISBN 978-9812771711, 2008.

[5]. Sladek, R. et al. (2007), *A genome-wide association study identifies novel risk loci for type 2 diabetes*. Nature, 445(7130): 881-885/