TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO CHUYÊN ĐỀ HỌC PHẦN**

**NGÔN NGỮ LẬP TRÌNH PYTHON**

**ĐỀ TÀI:**

**SỬ DỤNG NGÔN NGỮ PYTHON XÂY DỰNG ỨNG DỤNG DỰ ĐOÁN CHỈ SỐ HẠNH PHÚC BẰNG THUẬT TOÁN HỒI QUY TUYẾN TÍNH**

|  |  |
| --- | --- |
|  |  |
| |  |  |  | | --- | --- | --- | | **Sinh viên thực hiện** | **: TRẦN VIỆT LONG**  **NGUYỄN HỮU QUYỀN**  **ĐẶNG HỒNG VIỆT** | | | **Giảng viên hướng dẫn** | **: VŨ VĂN ĐỊNH** | | **Ngành** | **: CÔNG NGHỆ THÔNG TIN** | | **Chuyên ngành** | **: CÔNG NGHỆ PHẦN MỀM** | | **Lớp** | **: D13CNPM7** | | **Khóa** | **: D13** | |  |
|  |  |

***Hà Nội, tháng 11 năm 2021***

**PHIẾU CHẤM ĐIỂM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **Nội dung thực hiện** | **Điểm** | **Chữ ký** |
| 1 | Nguyễn Hữu Quyền |  |  |  |
| 2 | Trần Việt Long |  |  |  |
| 3 | Đặng Hồng Việt |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Họ và tên giảng viên** | **Chữ ký** | **Ghi chú** |
| Giảng viên chấm 1: |  |  |
| Giảng viên chấm 2: |  |  |

**MỤC LỤC**

Trang

[**CHƯƠNG 1. TỔNG QUAN VỀ NGÔN NGỮ PYTHON 1**](#_Toc88084529)

[1.1 Ngôn ngữ lập trình python. 1](#_Toc88084530)

[1.1.1 Sự hình thành. 1](#_Toc88084531)

[1.1.2 Lịch sử phát triển 2](#_Toc88084532)

[1.2 Ứng dụng của python 3](#_Toc88084533)

[1.2.1 Làm trang web với framework của python 3](#_Toc88084534)

[1.2.2Python dùng để tạo nguyên mẫu phần mềm 3](#_Toc88084535)

[1.2.3 Ứng dụng trong khoa học tính toán 3](#_Toc88084536)

[1.2.4 Viết tool để tự động công việc 3](#_Toc88084537)

[1.2.5 Khoa học máy tính 4](#_Toc88084538)

[1.2.6 Lĩnh vực IoT – Internet vạn vật 4](#_Toc88084539)

[1.2.7 Làm dự án Blockchain với python 4](#_Toc88084540)

[1.2.8 Làm game với PyGame 5](#_Toc88084541)

[1.2.9 Machine Learning 5](#_Toc88084542)

[1.2.10 Là ngôn ngữ được sử dụng cho mục đích giảng dạy 5](#_Toc88084543)

[1.3 Ưu nhược điểm của python 6](#_Toc88084544)

[1.3.1 Ưu điểm 6](#_Toc88084545)

[1.3.2 Nhược điểm 7](#_Toc88084546)

[**CHƯƠNG 2 XÂY DỰNG ỨNG DỤNG DỰ ĐOÁN CHỈ SỐ HẠNH PHÚC BẰNG THUẬT TOÁN HỒI QUY TUYẾN TÍNH** 10](#_Toc88084547)

[2.1 Thuật toán Linear Regression 10](#_Toc88084548)

[2.1.1 Giới thiệu 10](#_Toc88084549)

[2.1.2 Phân tích toán học 10](#_Toc88084551)

[2.1.2.1 Dạng của Linear Regression 10](#_Toc88084552)

[2.1.2.2 Sai số dự đoán 10](#_Toc88084553)

[2.1.2.3 Hàm mất mát 11](#_Toc88084557)

[2.1.2.4 Nghiệm cho bài toán Linear Regression 11](#_Toc88084565)

[2.2 Bài toán dự đoán suy tim 12](#_Toc88084576)

[2.2.1 Phát biểu bài toán 12](#_Toc88084577)

[2.2.2 Yêu cầu của bài toán 12](#_Toc88084578)

[2.2.3 Xây dựng bộ dữ liệu 13](#_Toc88084579)

[2.2.4 Áp dụng thuật toán vào bài toán 14](#_Toc88084580)

[**CHƯƠNG 3: CÀI ĐẶT** 17](#_Toc88084581)

[3.1 Cài đặt ngôn ngữ python 17](#_Toc88084582)

[3.2 Demo chương trình 20](#_Toc88084583)

[**KẾT LUẬN** 22](#_Toc88084585)

**DANH MỤC HÌNH ẢNH**

trang

[Hình 1:Guido cha đẻ của ngôn ngữ python 1](#_Toc88086350)

[Hình 2: Python giúp tạo tool để tự động hóa công việc. 4](#_Toc88086351)

[Hình 3: Lập trình python có liên hệ mật thiết với trí tuệ nhân tạo. 5](#_Toc88086352)

[Hình 7: Dữ liệu chuẩn đoán bệnh 14](#_Toc88086353)

[Hình 8: Download python. 17](#_Toc88086354)

[Hình 9: Cài đặt python. 17](#_Toc88086355)

[Hình 10: Chọn tính năng. 18](#_Toc88086356)

[Hình 11: Chọn tính năng nâng cao. 18](#_Toc88086357)

[Hình 12: Python đang được cài đặt. 19](#_Toc88086358)

[Hình 13: kết quả. 19](#_Toc88086359)

[Hình 14: Giao diện nhập dữ liệu. 20](#_Toc88086360)

[Hình 15: Giao diện hiển thị kết quả dự đoán. 21](#_Toc88086361)

**LỜI MỞ ĐẦU**

Trong cuộc sống của mỗi chúng ta khi nền kinh tế ngày càng phát triển, ngành công nghệ thông tin trở thành một trở thủ, điều không thể thiếu với hầu hết các ngành nghề. Có thể ban đầu nhiều người sẽ cảm thấy chưa cần thiết hoặc chưa sẵn sàng sử dụng nó vì nghĩ rằng phần mềm là một khái niệm gì đó rất mới mẻ, mới lạ, khó sử dụng.

Ngày nay trong cuộc sống 4.0 của chúng ta gần như không thể thiếu được những khái niệm liên quan đến công nghệ thông tin, từ các phương tiện truyền thông, xã hội cho đến lĩnh vực kinh doanh, quản lý,…Mọi thứ đều cần những phần mềm cũng như sản phẩm của công nghệ thông tin để hỗ trợ giúp nâng cao phát triển vững mạnh hơn.

Cùng với sự phát triển của nhiều ngôn ngữ lập trình như PHP, Java thì Python là một cái tên đáng chú ý. Hiện nay ngôn ngữ Python là một ngôn ngữ được sử dụng rất phổ biến. Python là một ngôn ngữ có hình thái sáng sửa, cấu trúc rõ ràng, thuận tiên cho người mới bắt đầu học lập trình. Cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu nhất. Python là ngôn ngữ lập trình đơn giản nhưng lại rất hiệu quả. Bên cạnh đó, Python là ngôn ngữ có tính hướng đối tượng cao.

Chính vì vậy mà em chọn đề tài “*Xây dựng ứng dụng dự đoán chỉ số hạnh phúc”* cho môn học này. Em rất mong nhận được sự ủng hộ của quý thầy, cô giáo để chúng em đạt được kết quả tốt nhất. Chúng em xin chân thành cảm ơn thầy Vũ Văn Định đã hướng dẫn, giúp đỡ em trong quá trình học tập. Cô đã giúp trang bị cho em kiến thức môn học và hơn cả là động lực để tiếp tục trên con đường học tập.

# CHƯƠNG 1. TỔNG QUAN VỀ NGÔN NGỮ PYTHON

## 1.1 Ngôn ngữ lập trình python.

### **1.1.1 Sự hình thành.**

Python là một [ngôn ngữ lập trình](https://vi.wikipedia.org/wiki/Ng%C3%B4n_ng%E1%BB%AF_l%E1%BA%ADp_tr%C3%ACnh) bậc cao cho các mục đích lập trình đa năng, do [Guido van Rossum](https://vi.wikipedia.org/w/index.php?title=Guido_van_Rossum&action=edit&redlink=1) tạo ra và lần đầu ra mắt vào năm [1991](https://vi.wikipedia.org/wiki/1990). Python được thiết kế với ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ. Python là ngôn ngữ có hình thức rất sáng sủa, cấu trúc rõ ràng, thuận tiện cho người mới học lập trình và là ngôn ngữ lập trình dễ học. được dùng rộng rãi trong phát triển [trí tuệ nhân tạo](https://vi.wikipedia.org/wiki/Tr%C3%AD_tu%E1%BB%87_nh%C3%A2n_t%E1%BA%A1o) hoặc [AI](https://vi.wikipedia.org/wiki/AI). Cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu.Vào tháng 7 năm 2018, Van Rossum đã từ chức lãnh đạo trong cộng đồng ngôn ngữ Python sau 30 năm làm việc.



Hình 1:Guido cha đẻ của ngôn ngữ python

Python hoàn toàn [tạo kiểu động](https://vi.wikipedia.org/w/index.php?title=T%E1%BA%A1o_ki%E1%BB%83u_%C4%91%E1%BB%99ng&action=edit&redlink=1) và dùng cơ chế [cấp phát bộ nhớ tự động](https://vi.wikipedia.org/wiki/Qu%E1%BA%A3n_l%C3%BD_b%E1%BB%99_nh%E1%BB%9B); do vậy nó tương tự như [Perl](https://vi.wikipedia.org/wiki/Perl), [Ruby](https://vi.wikipedia.org/wiki/Ruby_(ng%C3%B4n_ng%E1%BB%AF_l%E1%BA%ADp_tr%C3%ACnh)), [Scheme](https://vi.wikipedia.org/wiki/Scheme), [Smalltalk](https://vi.wikipedia.org/wiki/Smalltalk), và [Tcl](https://vi.wikipedia.org/wiki/Tcl). Python được phát triển trong một dự án mã mở, do tổ chức phi lợi nhuận Python Software Foundation quản lý.

Ban đầu, Python được phát triển để chạy trên nền [Unix](https://vi.wikipedia.org/wiki/Unix). Nhưng rồi theo thời gian, Python dần mở rộng sang mọi [hệ điều hành](https://vi.wikipedia.org/wiki/H%E1%BB%87_%C4%91i%E1%BB%81u_h%C3%A0nh) từ [MS-DOS](https://vi.wikipedia.org/wiki/MS-DOS) đến [Mac OS](https://vi.wikipedia.org/wiki/Mac_OS), [OS/2](https://vi.wikipedia.org/wiki/OS/2), [Windows](https://vi.wikipedia.org/wiki/Microsoft_Windows), [Linux](https://vi.wikipedia.org/wiki/Linux) và [các hệ điều hành khác thuộc họ Unix](https://vi.wikipedia.org/wiki/T%C6%B0%C6%A1ng_t%E1%BB%B1_Unix). Mặc dù sự phát triển của Python có sự đóng góp của rất nhiều cá nhân, nhưng Guido van Rossum hiện nay vẫn là tác giả chủ yếu của Python. Ông giữ vai trò chủ chốt trong việc quyết định hướng phát triển của Python.

Python luôn được xếp vào hạng những ngôn ngữ lập trình phổ biến nhất.

**1.1.2 Lịch sử phát triển**

Sự phát triển của ngôn ngữ lập trình Python phát triển theo các giai đoạn sau:

***Giai đoạn 1***: Là các bản phát hành python 1x. Giai đoạn này bắt đầu từ năm 1990 cho đến 2000. Năm 1990 đến 1995 , Guido van Rossum làm việc tại CWI trung tâm toán tin tại Amsterdam, Hà Lan. Do vậy phiên bản Python đầu tiên do trung tâm toán tin phát hành. Phiên bản cuối cùng tại CWI là python đời 1.2.

Năm 1995, ông chuyển sang CNRI ở Reston, Virginia. Ông đã nâng cấp đến Python 1.6 – đây cũng là bản cuối cùng của ông tại CNRI.

Sau đó ông rời CNRI và làm việc với các lập trình viên tự do chuyên viết phần mềm thương mại. Lúc đó ông nảy sinh ý tưởng kết hợp giữa Python với các phần mềm tuân theo GPL. Và chính CNRI và FSF hợp tác để làm bản quyền này của ông. Cũng trong năm này, ông được giải thưởng vì sự phát triển phần mềm tự do của FSF trao tặng.

Một thời gian sau phiên bản 16.1 ra đời tuân theo bản quyền GPL.

***Giai đoạn 2***: Năm 2000, Guido van Rossum cùng nhóm phát triển dời đến BeOpen.com để thành lập team phát triển Python. Và phiên bản 2.0 được phát triển tại đây. Sau đó Guido và các thành viên tham gia vào Digital Creations.

Sau đó phiên bản 2.1 ra đời dựa trên Python 1.6.1 và phiên bản 2.0. Từ thời điểm này Python thuộc sở hữu của PSF- đây là một tổ chức phi lợi nhuận.

***Giai đoạn 3***: Phiên bản 3x không tương thích hoàn toàn với 2x nhưng lại có công cụ hỗ trợ chuyển đổi từ phiên bản này sang phiên bản kia. Nguyên tắc chủ đạo trong việc phát triển Python ở phiên bản 3X chính là hạn chế sự trùng lặp về chức năng của Python. Ở phiên bản 3X có những thay đổi trong cú pháp và bổ sung thêm một số cú pháp mới.

**1.2 Ứng dụng của python**

**1.2.1 Làm trang web với framework của python**

Với ngôn ngữ Python sẽ giúp bạn làm trang web một cách đơn giản hơn.

Hiện nay những web frameworks Django và Flask ngày càng được sử dụng phổ biến nên việc phát triển web bằng Python dễ dàng hơn.

Từ khung ứng dụng web cho đến các phiên bản thiết kế đầy đủ các thư viện cần thiết thực sự hữu ích cho người dùng.

### 1.2.2Python dùng để tạo nguyên mẫu phần mềm

Đây là một đặc điểm lợi thế của ngôn ngữ Python. Chỉ với Python bạn sẽ có thể tạo ra những nguyên mẫu phần mềm – bản chạy thử. Tuy nhiên tốc độ của Python sẽ chậm hơn so với Java hay C++. Do vậy bạn cần cân nhắc về nguồn lực cũng như hiệu quả để lựa chọn ngôn ngữ phù hợp.

### 1.2.3 Ứng dụng trong khoa học tính toán

Ngoài các ứng dụng trên, lập trình python còn có ứng dụng nào khác không? Đó chính là sự ứng dụng vào khoa học và tính toán. Ngôn ngữ Python sở hữu thư viện dành cho lĩnh vực  khoa học, tính toán số liệu như NumPy, SciPy. Ngoài ra, Python còn có  EarthLy dành cho khoa học trái đất, AstroPy cho các ngành thiên văn học, Biopython cho ngành sinh học, Graph-tool cho ngành phân tích biểu đồ, … Ngôn ngữ lập trình Python được sử dụng trọng khai thác dữ liệu,…

### 1.2.4 Viết tool để tự động công việc

Thông thường, dân IT luôn tự viết tool để công việc trở nên dễ dàng hơn. Thay vì sử dụng ngôn ngữ Java, PHP để viết tool thì việc sử dụng ngôn ngữ lập trình Python sẽ giúp lập trình viên chủ động hơn trong công việc. Bởi tốc độ viết tool bằng ngôn ngữ này nhanh gấp 10 lần so với các ngôn ngữ khác.

Bạn có thể tạo ra từ điển, quét trang web, làm việc với các tệp khác,….



Hình 2: Python giúp tạo tool để tự động hóa công việc.

### 1.2.5 Khoa học máy tính

Trong Python có rất nhiều thư viện quan trọng trong khoa học máy tính như:

* Open Computer Vision viết tắt là Opencv – một thư viện mã nguồn mở cho xử lý về thị giác máy tính, xử lý ảnh và machine learning.
* Numpys: là thư viện hỗ trợ cho việc tính toán các mảng đa dạng. Numpy rất hữu ích với những hàm liên quan đến Đại Số Tuyến Tính.
* Scipy: Là phần mềm nguồn mở dành cho lĩnh vực toán học, khoa học và kỹ thuật. Scipy gồm các tập con cho đại số tuyến tính, tích hợp và thống kê.
* Pandas: là một thư viện dành cho công việc phân tích dữ liệu.

### 1.2.6 Lĩnh vực IoT – Internet vạn vật

Với các nền tảng nhúng, hoặc những ứng dụng cần truy xuất dữ liệu lớn thì Python là giải pháp phù hợp. Python còn được ứng dụng trên trang instagram hay Pinterest nên ngày càng quen thuộc với người dùng.

### 1.2.7 Làm dự án Blockchain với python

Với những dự án liên quan đến Blockchain, đòi hỏi bạn cần phải biết cách làm  việc với HTTP Clients và thư viện cần thiết. Thông qua Flask Framework, người dùng sẽ sử dụng HTTP requests để dễ dàng liên lạc với blockchain của mình qua Internet.

### 1.2.8 Làm game với PyGame

Pygame nằm trong thư viện SDL sẽ giúp việc phát triển các game 2D một cách dễ dàng hơn bao giờ hết.

### 1.2.9 Machine Learning

Nắm vững ngôn ngữ lập trình Python sẽ giúp các lĩnh vực liên quan đến Machine Learning, Ai một cách dễ dàng. Bởi Python có những hệ thống dữ liệu quan trọng như: SciPy, scikit-learn, panda,…

****

Hình 3: Lập trình python có liên hệ mật thiết với trí tuệ nhân tạo.

### 1.2.10 Là ngôn ngữ được sử dụng cho mục đích giảng dạy

Với nhiều ứng dụng trong các lĩnh vực nên ngôn ngữ này được sử dụng cho mục đích giảng dạy như: phân tích dữ liệu, tự động hóa, lập trình game 2D,…

Với những thông tin trên đây, hy vọng người đọc đã có thể giải đáp những thắc mắc của riêng mình về lập trình giao diện python. Và để công việc thích ứng được với nhu cầu của thời đại, thì lập trình viên ngoài am hiểu các ngôn ngữ lập trình thì cần phải biết về mô hình Agile. Đây là mô hình được nhiều doanh nghiệp hàng đầu trên thế giới ứng dụng như Facebook, Apple, Amazon,…Với những lợi thế như tính linh hoạt, sáng tạo, hiệu quả; mô hình Agile sẽ đem đến việc phát triển dự án một cách thông minh nhất.

**1.3 Ưu nhược điểm của python**

**1.3.1 Ưu điểm**

Với cộng đồng python mới nổi và các thư viện mã nguồn mở, python đã phát triển thành một gói phát triển phần mềm hoàn chỉnh.

* **Được thúc đẩy bởi cộng đồng rộng lớn và năng động**

Python có một trong những cộng đồng tích cực và được biết đến nhiều nhất giúp họ cải thiện liên tục con trăn. Không có gì ngạc nhiên khi nó là nền tảng được xếp hạng hàng đầu về tràn ngăn xếp.  
Ngôn ngữ Python được phân phối theo giấy phép nguồn mở giúp phát triển dễ dàng thông qua các đóng góp nguồn mở.

* **Đường cong học tập**

Với hầu hết các ngôn ngữ lập trình, đường cong học tập của họ có xu hướng phát triển theo hình parabol theo thời gian, điều đó có nghĩa là khó có thể nắm bắt sớm nhưng ngay khi bạn làm quen với ngôn ngữ này, việc học sẽ trở nên dễ dàng. Nhưng trong trường hợp python, việc học rất dễ dàng vì cú pháp dễ dàng và viết tay ngắn.

* **Thư viện bên thứ ba**

Trình cài đặt gói python chuẩn (PIP) có thể giúp bạn cài đặt nhiều mô-đun làm cho python tương tác. Các thư viện và mô-đun này có thể tương tác từ các giao thức internet, lệnh gọi hệ điều hành và nhiều thứ khác.  
Bạn có thể lập trình socket cho mạng và sử dụng mô-đun os cho các lệnh gọi hệ thống chạy các luồng cấp người dùng.

* **Tích hợp với các ngôn ngữ khác**

Các thư viện tích hợp như Cython và Jython làm cho python tích hợp với c / c ++ và java để phát triển đa nền tảng. Điều này làm cho python thậm chí còn mạnh mẽ hơn vì tất cả chúng ta đều biết không có ngôn ngữ nào là hoàn chỉnh và bạn nên sử dụng một ngôn ngữ duy nhất để phát triển mọi thứ. Mỗi ngôn ngữ đều có sở trường riêng, vì vậy sử dụng các thư viện này, bạn có thể tận hưởng các tính năng mạnh mẽ của từng ngôn ngữ.

* **Năng xuất**

Với triết lý bao gồm pin python, các nhà phát triển sẽ có một khởi đầu thuận lợi mà không cần tải xuống các mô-đun và gói riêng biệt. Cùng với cú pháp dễ dàng của python và tạo mẫu nhanh, năng suất tăng gần 40% so với các ngôn ngữ lập trình truyền thống.

So với các ngôn ngữ lập trình khác, Python được các nhà phát triển sử dụng rộng rãi nhất hiện nay. Các lợi ích quan trọng của ngôn ngữ Python là nó dễ đọc và mượt mà để kiểm tra và học hỏi. Việc cài đặt các gói và viết một phần mềm bằng Python ít phức tạp hơn so với C hoặc C ++. Một số lợi ích khác của lập trình Python là không có vi rút máy tính nào có thể gây ra lỗi phân đoạn vì không có khái niệm về con trỏ hoặc tham chiếu trong python. Một lợi thế quan trọng của ngôn ngữ Python so với các ngôn ngữ lập trình truyền thống là nó có khả năng ứng dụng và chấp nhận rộng rãi, đồng thời được các nhà khoa học, kỹ sư và nhà toán học sử dụng một cách đáng kể. do đó Python rất có lợi cho việc tạo mẫu và tất cả các loại thử nghiệm. Nó cũng được sử dụng cùng lúc với việc tạo hoạt ảnh cho phim và trong học máy và học sâu.

**1.3.2 Nhược điểm**

Chúng ta đã thấy những ưu điểm chính của ngôn ngữ lập trình phổ biến Python. Nhưng chúng ta đều biết có hai mặt của một đồng xu!  
Python thực sự cũng có một số nhược điểm khiến các nhà phát triển tránh xa nó. Vì vậy, chúng ta hãy xem từng cái một:

* **Tốc độ chậm**

Python sử dụng trình thông dịch tải từng dòng một thay vì trình biên dịch thực thi toàn bộ tệp cùng một lúc. Điều này làm cho quá trình biên dịch chậm hơn và có xu hướng hoạt động chậm. Đây là lý do chính mà các lập trình viên cạnh tranh không sử dụng python. C ++ cung cấp nhiều tính toán hơn trên giây thay vì python. Hơn nữa, đây là lý do tại sao python không được sử dụng rộng rãi trong phát triển ứng dụng.

* **Phát hiện lỗi trong mã**

Vì python sử dụng trình thông dịch nên lỗi trong mã không xuất hiện trong quá trình biên dịch mã. Python sử dụng các biến được nhập động làm cho việc thử nghiệm trở nên bận rộn hơn. Tất cả những lỗi này đều là lỗi thời gian chạy mà không nhà phát triển nào muốn trong chương trình của họ. Điều này làm cho python ít được sử dụng hơn cho các phát triển ứng dụng GUI.

* **Yếu trong thiết bị di động**

Chúng tôi đã thấy python trong máy chủ web và ứng dụng máy tính để bàn cùng với các tập lệnh mà nó được sử dụng. Nhưng nó không được coi là thuận lợi cho các thiết bị di động vì nó sử dụng nhiều bộ nhớ hơn và xử lý chậm so với các ngôn ngữ khác.

* **Tiêu thụ bộ nhớ lớn**

Cấu trúc thiết kế Python là do nó sử dụng bộ nhớ lớn trong khi xử lý so với các ngôn ngữ khác như C / C ++. Điều này làm cho Python hoàn toàn không cho sự phát triển trong điều kiện hạn chế bộ nhớ chặt chẽ.

Ngôn ngữ này được coi là một nền tảng ít phù hợp hơn để phát triển di động và phát triển trò chơi. Nó thường được sử dụng trên máy tính để bàn và máy chủ, nhưng có một số gói di động tốt nhất được sử dụng chủ yếu với Python. Mọi nhược điểm khác của Python là lỗi thời gian chạy. Ngôn ngữ này có rất nhiều hạn chế về thiết kế .Python thực thi với trình thông dịch thay vì trình biên dịch, điều này làm giảm tốc độ hiệu suất.

**CHƯƠNG 2 XÂY DỰNG ỨNG DỤNG DỰ ĐOÁN CHỈ SỐ HẠNH PHÚC BẰNG THUẬT TOÁN HỒI QUY TUYẾN TÍNH**

# 2.1 Thuật toán Linear Regression

**2.1.1 Giới thiệu**

"Hồi quy tuyến tính" là một phương pháp thống kê để hồi quy dữ liệu với biến phụ thuộc có giá trị liên tục trong khi các biến độc lập có thể có một trong hai giá trị liên tục hoặc là giá trị phân loại. Nói cách khác "Hồi quy tuyến tính" là một phương pháp để dự đoán biến phụ thuộc (Y) dựa trên giá trị của biến độc lập (X). Nó có thể được sử dụng cho các trường hợp chúng ta muốn dự đoán một số lượng liên tục. Ví dụ, dự đoán giao thông ở một cửa hàng bán lẻ, dự đoán thời gian người dùng dừng lại một trang nào đó hoặc số trang đã truy cập vào một website nào đó v.v...

**2.1.2 Phân tích toán học**

### **2.1.2.1 Dạng của Linear Regression**

Trong phương trình (1) phía trên, nếu chúng ta đặt w = [ w 0 , w 1 , w 2 , w 3 ] T = là vector (cột) hệ số cần phải tối ưu và ¯ x = [ 1 , x 1 , x 2 , x 3 ] (đọc là x bar trong tiếng Anh) là vector (hàng) dữ liệu đầu vào mở rộng. Số 1 ở đầu được thêm vào để phép tính đơn giản hơn và thuận tiện cho việc tính toán. Khi đó, phương trình (1) có thể được viết lại dưới dạng: y ≈ ¯ x w = ^ y

Chú ý rằng ¯ x là một vector hàng.

### **2.1.2.2 Sai số dự đoán**

### Chúng ta mong muốn rằng sự sai khác e giữa giá trị thực y và giá trị dự đoán ^ y (đọc là y hat trong tiếng Anh) là nhỏ nhất. Nói cách khác, chúng ta muốn giá trị sau đây càng nhỏ càng tốt:

### 1/2e2=12(y−^y)2=12(y−¯xw)2

### trong đó hệ số 1/2 (lại) là để thuận tiện cho việc tính toán (khi tính đạo hàm thì số 1/2 sẽ bị triệt tiêu). Chúng ta cần e2 vì e = y − ^ y có thể là một số âm, việc nói e nhỏ nhất sẽ không đúng vì khi e = − ∞ là rất nhỏ nhưng sự sai lệch là rất lớn.

### **2.1.2.3 Hàm mất mát**

### Điều tương tự xảy ra với tất cả các cặp (input, outcome )( x i , y i ),i = 1 , 2 , … , N , với N là số lượng dữ liệu quan sát được. Điều chúng ta muốn, tổng sai số là nhỏ nhất, tương đương với việc tìm w để hàm số sau đạt giá trị nhỏ nhất:

### L(w)=12N∑i=1(yi−¯xiw)2 (2)

### Hàm số L(w) được gọi là hàm mất mát (loss function) của bài toán Linear Regression. Chúng ta luôn mong muốn rằng sự mất mát (sai số) là nhỏ nhất, điều đó đồng nghĩa với việc tìm vector hệ số w sao cho giá trị của hàm mất mát này càng nhỏ càng tốt. Giá trị của w làm cho hàm mất mát đạt giá trị nhỏ nhất được gọi là điểm tối ưu (optimal point), ký hiệu:

### w∗=argminwL(w)

### Trước khi đi tìm lời giải, chúng ta đơn giản hóa phép toán trong phương trình hàm mất mát (2). Đặt y = [ y 1; y 2;… ; y N ] là một vector cột chứa tất cả các output của training data; X = [ ¯ x 1 ; ¯ x 2 ; … ; ¯ x N ] là ma trận dữ liệu đầu vào (mở rộng) mà mỗi hàng của nó là một điểm dữ liệu. Khi đó hàm số mất mát L(w) được viết dưới dạng ma trận đơn giản hơn:

### L(w)=12N∑i=1(yi−¯xiw)2L(w)=12∑i=1N(yi−x¯iw)2=12∥y−¯Xw∥22 (3)

### với ∥ z ∥ 2 là Euclidean norm (chuẩn Euclid, hay khoảng cách Euclid), nói cách khác ∥ z ∥ 2 2 là tổng của bình phương mỗi phần tử của vector z . Tới đây, ta đã có một dạng đơn giản của hàm mất mát được viết như phương trình (3).

### **2.1.2.4 Nghiệm cho bài toán Linear Regression**

### Cách phổ biến nhất để tìm nghiệm cho một bài toán tối ưu (chúng ta đã biết từ khi học cấp 3) là giải phương trình đạo hàm (gradient) bằng 0! Tất nhiên đó là khi việc tính đạo hàm và việc giải phương trình đạo hàm bằng 0 không quá phức tạp. Thật may mắn, với các mô hình tuyến tính, hai việc này là khả thi. Đạo hàm theo w của hàm mất mát là:

### ∂ L (w) ∂ w = ¯ X T (¯ X w − y )

### Các bạn có thể tham khảo bảng đạo hàm theo vector hoặc ma trận của một hàm số trong mục D.2 của tài liệu này. Đến đây tôi xin quay lại câu hỏi ở phần Sai số dự đoán phía trên về việc tại sao không dùng trị tuyệt đối mà lại dùng bình phương. Câu trả lời là hàm bình phương có đạo hàm tại mọi nơi, trong khi hàm trị tuyệt đối thì không (đạo hàm không xác định tại 0).

### Phương trình đạo hàm bằng 0 tương đương với:

### ¯ X T ¯ X w = ¯ X T y ≜ b (4)

### (ký hiệu ¯ X T y ≜ b nghĩa là đặt ¯ X T y bằng b).

### Nếu ma trận vuông A ≜ ¯ X T ¯ X khả nghịch (non-singular hay invertible) thì phương trình (4) có nghiệm duy nhất: w = A − 1 b.

### Vậy nếu ma trận A không khả nghịch (có định thức bằng 0) thì sao? Nếu các bạn vẫn nhớ các kiến thức về hệ phương trình tuyến tính, trong trường hợp này thì hoặc phương trinh (4) vô nghiệm, hoặc là nó có vô số nghiệm. Khi đó, chúng ta sử dụng khái niệm giả nghịch đảo A † (đọc là A dagger trong tiếng Anh). (Giả nghịch đảo (pseudo inverse) là trường hợp tổng quát của nghịch đảo khi ma trận không khả nghịch hoặc thậm chí không vuông. Trong khuôn khổ bài viết này, tôi xin phép được lược bỏ phần này, nếu các bạn thực sự quan tâm, tôi sẽ viết một bài khác chỉ nói về giả nghịch đảo. Xem thêm: Least Squares, Pseudo-Inverses, PCA & SVD.)

### Với khái niệm giả nghịch đảo, điểm tối ưu của bài toán Linear Regression có dạng:

### w = A † b = (¯ X T ¯ X) † ¯ X T y (5)

**2.2 Bài toán dự đoán chỉ số hạnh phúc**

**2.2.1 Phát biểu bài toán**

Bài toán chuẩn đoán bệnh nhân suy tim đưa ra tất cả các thông tin của một người có khả năng mắc bệnh để từ đó làm căn cứ chuẩn đoán khả năng bị bệnh của người đó ở thời điểm hiện tại cho đến thời điểm sau đó.

- Giá trị input: Thông tin, đặc tính của một người có có khả năng mắc bệnh hoặc không.

- Giá trị output: Kết quả người được chuẩn đoán có bị bệnh hay không.

### 2.2.2 Yêu cầu của bài toán

- Lấy dữ liệu về mô tả đặc tính của con người.

- Trích chọn đặc trưng từ tập dữ liệu lấy được .

- Tiến hành huấn luyện tập dữ liệu .

- Chuẩn đoán bệnh của người .

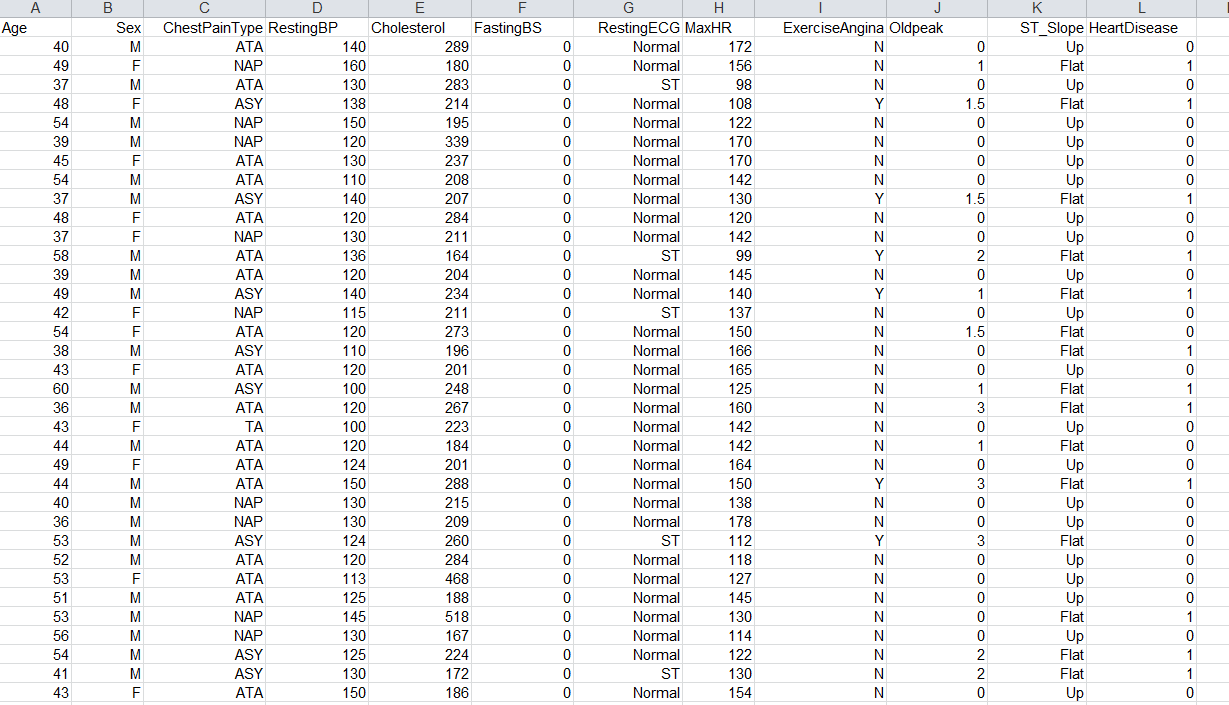
### 2.2.3 Xây dựng bộ dữ liệu

Giả sử một bệnh viện có một tập dữ liệu lớn gồm 3000 bệnh nhân đang được chuẩn đoán khác nhau, sau một thời gian chuẩn đoán thì xác định được 100 người mắc bệnh, 2900 người không mắc bệnh. Bệnh viện sẽ tổng hợp toàn bộ dữ liệu và phân loại bệnh nhân theo trường hợp mắc bệnh hoặc không mắc bệnh.

Với các thuộc tính sau:

* ***Age***: Tuổi
* ***Sex***: giới tính “M”-“nam”,”F”-“nữ”
* ***ChestPainType***:  loại đau ngực [TA: Đau thắt ngực điển hình, ATA: Đau thắt ngực không điển hình, NAP: Đau thắt ngực không, ASY: Không có triệu chứng].
* ***RestingBP***: huyết áp khi nghỉ ngơi [mm Hg].
* ***Cholesterol***: cholesterol huyết thanh [mm / dl].
* ***FastingBS***: đường huyết lúc đói [1: nếu FastingBS> 120 mg / dl, 0: nếu không].
* ***RestingECG***: kết quả điện tâm đồ lúc nghỉ [Bình thường: Bình thường, ST: có bất thường sóng ST-T (đảo ngược sóng T và / hoặc ST chênh lên hoặc xuống> 0,05 mV), LVH: cho thấy phì đại thất trái có thể hoặc xác định theo tiêu chuẩn của Estes].
* ***MaxHR:*** nhịp tim tối đa đạt được [Giá trị số từ 60 đến 202].
* ***ExerciseAngina:*** đau thắt ngực do tập thể dục [Y: Có, N: Không].
* ***Oldpeak:*** Độ trầm cảm.
* ***ST\_Slope:*** Độ dốc kết quả điện tâm đồ [Lên: dốc lên, Flat: bằng phẳng, Xuống: dốc xuống].
* ***HeartDisease***: lớp đầu ra [1: bệnh tim, 0: Bình thường].

Nhân viên lấy mẫu dữ liệu các bệnh nhân đã được phân loại chuẩn đoán mắc bệnh để phân tích cho bệnh nhân mới nhập viện.



Hình 7: Dữ liệu chuẩn đoán bệnh

### 2.2.4 Áp dụng thuật toán vào bài toán

Sử dụng thuật toán KNN để chuẩn đoán suy tim

Dữ liệu chuẩn đoán :

* Age
* Sex
* ChestPainType
* RestingBP
* Cholesterol
* FastingBS
* RestingECG
* MaxHR
* ExerciseAngina
* Oldpeak
* ST\_Slope

=>> Có mắc bệnh hay không ?? Outcome là 1 hay 0.

Dùng thuật toán KNN để phân loại rằng bệnh nhân được chuẩn đoán đó có bị suy tim hay không.

# 

# CHƯƠNG 3: CÀI ĐẶT

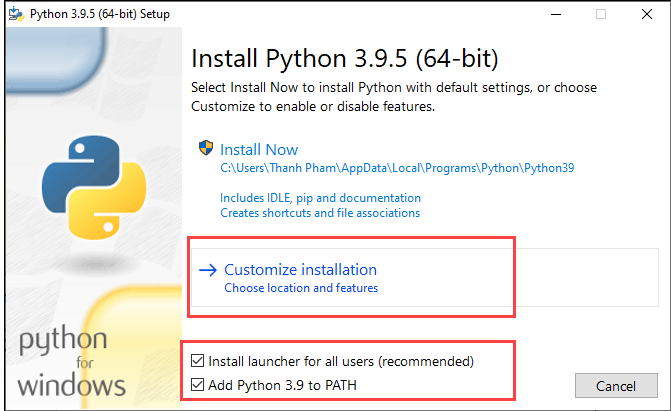
## 3.1 Cài đặt ngôn ngữ python

Truy cập liên kết <https://www.python.org/downloads/> để tải xuống bản phát hành mới nhất của Python. Bản hiện tại là Python 3.9.5 cho hệ điều hành Windows.

****

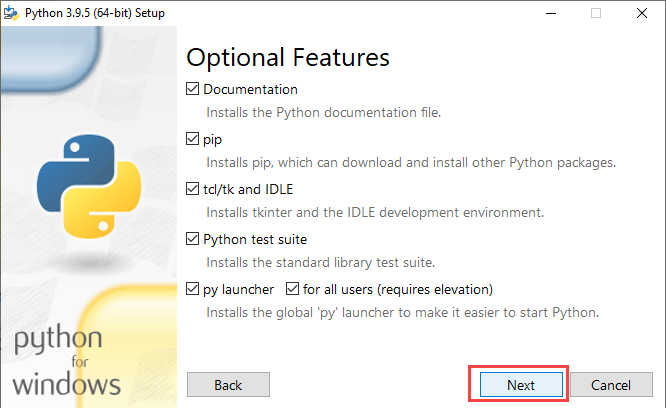
Hình 8: Download python.

Click đúp chuột vào file vừa tải xuống để tiến hành cài đặt Python:



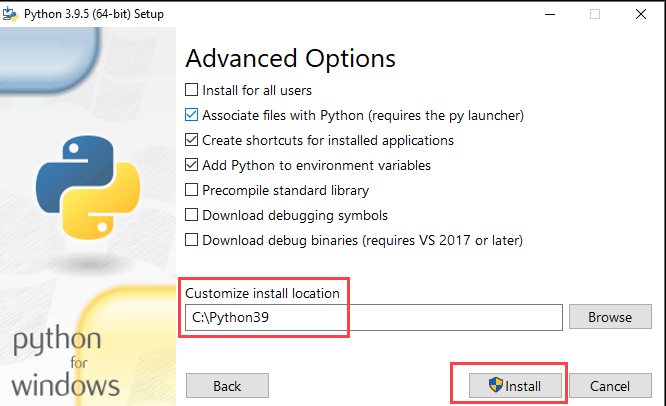
Hình 9: Cài đặt python.

Chọn **Customize installation**:



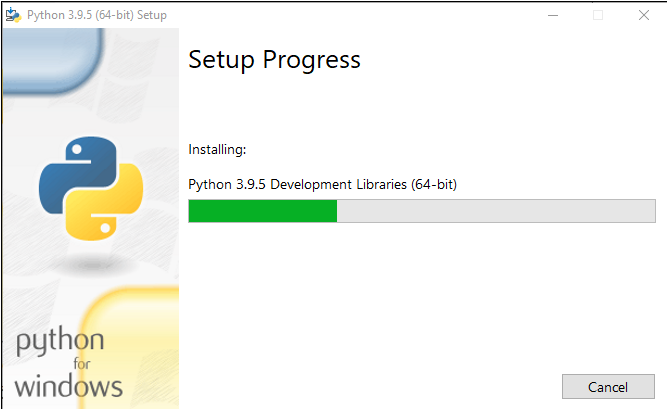
Hình 10: Chọn tính năng.

Chọn tất cả các tính năng tùy chọn, click Next:



Hình 11: Chọn tính năng nâng cao.

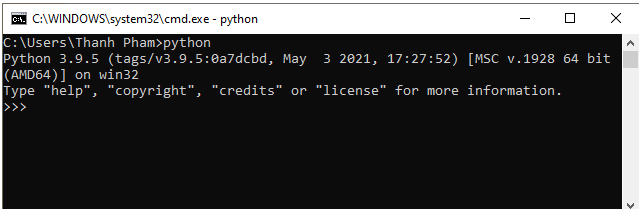
Chọn các tính năng tùy chọn như trong hình, click Install:



Hình 12: Python đang được cài đặt.

Python 3.9.5 đang được cài đặt:

Giả sử máy tính của bạn đã khởi động xong, bây giờ, hãy thử chạy python trên command line (CMD). Gõ lệnh python trong trường hợp python2 hoặc python3 trong trường hợp python3. Kết quả.

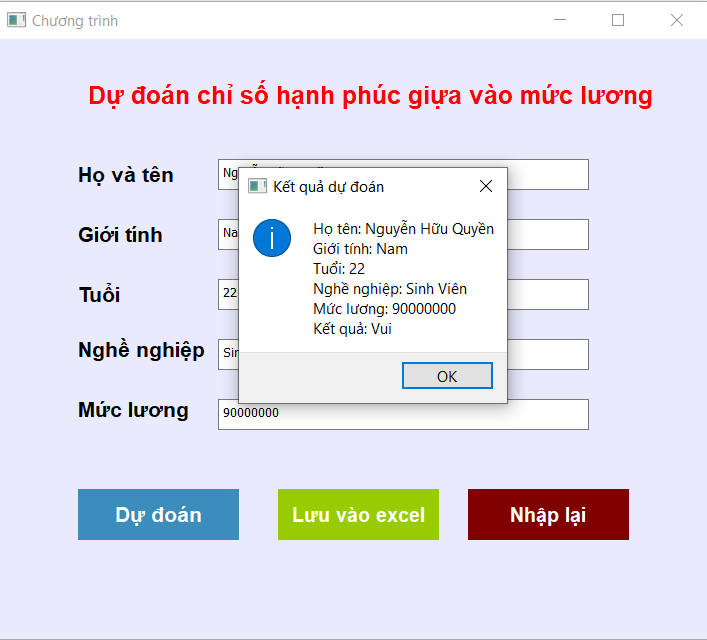


Hình 13: kết quả.

## 3.2 Demo chương trình



Hình 14: Giao diện nhập dữ liệu.

****

Hình 15: Giao diện hiển thị kết quả dự đoán.

### Giao diện kết quả dự đoán cho biết 1 số thông tin như tên bệnh nhân chuẩn đoán, số điện thoại liên hệ, kết quả chuẩn đoán và đọ chính xác.

# KẾT LUẬN

Với kiến thức nền tảng đã được học ở trường và bằng sự nỗ lực của mình, em đã hoàn thành đề tài *“Xây dựng ứng dụng dự đoán chỉ số hạnh phúc”* Mặc dù đã cố gắng và đầu tư rất nhiều nhưng do thời gian có hạn, kiến thức còn non yếu nên phần mềm chắc chắn còn nhiều hạn chế. Em rất mong nhận được sự thông cảm và góp ý của thầy, cô giáo để đề tài của chúng em được hoàn thiện hơn.

Một lần nữa chúng em xin chân thành cảm ơn Thầy Vũ Văn Định đã tận tình giúp đỡ em trong suốt thời gian thực hiện đề tài này. Em xin chân thành cảm ơn cô.