### BỘ GIÁO DỤC VÀ ĐÀO TẠO TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



KHOA CÔNG NGHỆ THÔNG TIN

**TRỰC QUAN HÓA DỮ LIỆU**

**BÁO CÁO BÀI TẬP MÔN HỌC BÀI TẬP: DV-TW-Lab01**

**Giảng viên hướng dẫn** :Lê Nhựt Nam

**Lớp** :CQ2022/4

### **Nhóm sinh viên thực hiện** : Đinh Viết Lợi- 22120188 Nguyễn Trần Lợi- 22120190 Nguyễn Nhật Long-22120194 Trần Minh Tâm- 22120323

*Hồ Chí Minh, ngày 29 tháng 10 năm 2024*

# MỤC LỤC

Mục lục

[PHẦN 1: BÁO CÁO NHÓM 2](#_bookmark0)

1. [Thông tin thành viên 2](#_bookmark1)
2. [Phân chia công việc 2](#_bookmark2)

[PHẦN 2: TỔNG QUAN ĐỒ ÁN 4](#_bookmark3)

1. [Yêu cầu đồ án 4](#_bookmark4)
   1. [Giới thiệu đồ án 4](#_bookmark5)
   2. [Yêu cầu đồ án 4](#_bookmark6)
2. [Mức độ hoàn thành 4](#_bookmark7)

[PHẦN 3: BÁO CÁO KẾT QUẢ 6](#_bookmark8)

1. [Thu thập dữ liệu 6](#_bookmark9)
2. [Khám phá dữ liệu 6](#_bookmark10)
3. [Trực quan và ý nghĩa 8](#_bookmark11)
   1. [Trực quan dữ liệu 8](#_bookmark12)
   2. [Mô hình dự đoán 9](#_bookmark13)

[PHẦN 4: TÀI LIỆU THAM KHẢO 13](#_bookmark14)

1. [Tài liệu nhóm 13](#_bookmark15)
2. [Tài liệu tham khảo 13](#_bookmark16)

# PHẦN 1: BÁO CÁO NHÓM

## Thông tin thành viên

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ và tên | Email | Vai trò chính |
| 22120188 | Đinh Viết Lợi | [22120188@student.hcmus.edu.vn](mailto:22120188@student.hcmus.edu.vn) | Trưởng nhóm |
| 22120190 | Nguyễn Trần Lợi | [22120190@student.hcmus.edu.vn](mailto:22120190@student.hcmus.edu.vn) |  |
| 22120194 | Nguyễn Nhật Long | [22120194@student.hcmus.edu.vn](mailto:22120194@student.hcmus.edu.vn) |  |
| 22120323 | Trần Minh Tâm | [22120323@student.hcmus.edu.vn](mailto:22120323@student.hcmus.edu.vn) |  |

## Phân chia công việc

Đồ án được xây dựng và hoàn thành trong khoảng thời gian từ : 20/10/2024 đến 1/11/2024.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nội dung công việc | Người phụ trách | Thời gian bắt đầu | Thời gian kết thúc | Kết quả mong muốn |
| Tìm nguồn dữ liệu | Tất cả thành viên | 20/10 | 23/10 | Chọn ra dataset phù hợp nhất |
| Quản lý, điều hành nhóm | Đinh Viết Lợi | 20/10 | 31/10 | Giúp các thành viên phối hợp tốt, hoàn thành công việc đúng yêu cầu và thời hạn. |
| Vẽ biểu đồ Radar, Line, Bar Chart | Viết Lợi | 20/10 | 3/11 | Vẽ biểu đồ đẹp, rút ra được ý nghĩa y học quan trọng. |
| Vẽ biểu đồ Heat map, Miền, Pie Chart | Trần Lợi | 24/10 | 29/10 | Vẽ biểu đồ đẹp, rút ra được ý nghĩa y học quan trọng. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vẽ biểu đồ Box Plot, Histogram, Heatmap | Trần Minh Tâm | 24/10 | 29/10 | Vẽ biểu đồ đẹp, rút ra được ý nghĩa y học quan trọng. |
| Vẽ biểu đồ Line, Violin, Venn | Nguyễn Nhật Long | 24/10 | 29/10 | Vẽ biểu đồ đẹp, rút ra được ý nghĩa y học quan trọng. |
| Mô hình học máy | Nguyễn Nhật Long | 27/10 | 1/11 | Dùng mô hình học máy đưa ra được dự đoán. |
| Làm báo cáo phần 3 | Đinh Viết Lợi | 1/11 | 2/11 | Báo cáo đầy đủ công việc của nhóm, đúng yêu cầu của giảng viên |
| Làm báo cáo phần 1,2,4 | Nguyễn Nhật Long | 1/11 | 2/11 | Báo cáo đầy đủ công việc của nhóm, đúng yêu cầu của giảng viên |

# PHẦN 2: TỔNG QUAN ĐỒ ÁN

## Yêu cầu đồ án

## Giới thiệu đồ án

* + Đồ án được thực hiện dựa trên yêu cầu của đồ án DV–TW– Lab01 thuộc lớp “Trực quan hóa dữ liệu - CQ2022/4” trường đại học Khoa Học Tự Nhiên thuộc Đại học Quốc gia TP.HCM học kỳ I năm học 2024-2025.
  + Đồ án nhằm trang bị cho sinh viên kiến thức về trực quan hóa dữ liệu bằng Python. Bằng cách học và tận dụng các thư viện Python như Matplotlib, Seaborn và Pandas để tạo ra nhiều loại biểu đồ tĩnh, tương tác và hoạt họa, sinh viên có thể khám phá các mối quan hệ ẩn trong bộ dữ liệu mà bạn đang tập trung vào trong một lĩnh vực cụ thể.
  + Kết quả của đồ án là các mã nguồn được chú thích đầy đủ tên biểu đồ, ý nghĩa, kết luận rút ra từ dữ liệu, có ý nghĩa thực tế trong việc giúp các chuyên gia, bệnh nhân trong công tác phòng tránh, điều trị bệnh tật được nằm trong thư mục D –TW –Lab01

## Yêu cầu đồ án

* + Bài làm không có có báo cáo sẽ không được chấm điểm.
  + Các thành viên không đóng góp cho dự án sẽ không nhận được điểm.
  + Các nguồn tham khảo (nếu có) cần được ghi đầy đủ trong báo cáo ở phần Tài liệu tham khảo. Lưu ý rằng cần phân biệt rõ giữa việc tham khảo và đạo văn.
  + Cá nhân hoặc nhóm nào vi phạm gian lận và không trung thực sẽ nhận 0 điểm trong khóa học.
  + Đối với buổi thực hành này, có một số hạn chế cụ thể mà chúng tôi yêu cầu bạn tuân thủ nghiêm ngặt:
  + Không sử dụng các giải pháp phần mềm nâng cao như Tableau cho mục đích trực quan hóa dữ liệu.
  + Nếu muốn sử dụng các thư viện bổ sung, vui lòng tham khảo ý kiến giảng viên trước để nhận được sự chấp thuận.
  + Trong khi việc áp dụng các thuật toán máy học đơn giản có thể cung cấp những hiểu biết sâu sắc hơn về dữ liệu của bạn, đây là yêu cầu tùy chọn và không phải là yêu cầu bắt buộc.

## Mức độ hoàn thành

* **Tổng quan**: nhóm hoàn thành đầy đủ các nội dung cơ bản được yêu cầu và thực hiện tìm hiểu vào các phần nội dung được khuyến khích. Bài làm được đầu tư cẩn thận về chất lượng và số lượng, báo cáo tổng hợp được tất cả quá trình làm việc và kết quả đạt được của nhóm.
* Bộ câu hỏi được chuẩn bị có tính phổ biến và ứng dụng cao, phù hợp với nhu cầu phân tích.
* Nhóm sử dụng đa dạng các loại biểu đồ để trực quan dữ liệu nhằm giải đáp các vấn đề được nêu ra. Tất cả thông tin cần thiết cho việc nắm bắt biểu đồ đều được trình bày rõ ràng, đầy đủ.

 Mức độ hoàn thành: 100%

* **Hạn chế:** nguồn dữ liệu được sử dụng chưa thật sự phù hợp, không đảm bảo hoàn toàn chất lượng.

# PHẦN 3: BÁO CÁO KẾT QUẢ

## Thu thập dữ liệu

* + Nguồn dữ liệu: Kaggle- [Heart Attack Risk Prediction Dataset](https://www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset/data).
  + Nguy cơ nhồi máu cơ tim hiện nay đang trở nên phổ biến hơn bao giờ hết khi giờ không chỉ người lớn tuổi đứng trước nguy cơ này mà ngay cả người trẻ từ mọi quốc gia, giới tính đều tiềm tàng khả năng này. Nhóm muốn tìm hiểu xem liệu nguyên nhân gì khiến cho việc căn bệnh này trở nên phổ biến như vậy, không chỉ tìm hiểu hết tất cả các nguyên nhân, thói quen tưởng chừng vô hại hằng ngày mà đồng thời từ đó tìm ra một lối sống lành mạnh để giảm nguy cơ trở thành nạn nhân của căn bệnh này.
  + “Bộ dữ liệu này là một sản phẩm tổng hợp được tạo ra bởi ChatGPT nhằm mô phỏng một trải nghiệm thực tế. Mục đích là cung cấp nền tảng cho người mới bắt đầu và những người đam mê dữ liệu, cho phép họ thực hành, học hỏi và khám phá từ một bộ dữ liệu phản ánh các tình huống trong thế giới thực” – Theo lời tác giả của bộ dữ liệu. Mặc dù là bộ dữ liệu là một sản phẩm tổng hợp tuy nhiên số lượng và chất lượng của bộ dữ liệu có thể tin cậy được khi so sánh với thực tế. Nhóm thu thập nguồn dữ liệu này tới Kaggle- một diễn đàn nơi mọi người chia sẻ các tập dữ liệu của mình.
  + Nguồn dữ liệu được sử dụng cho mục đích phân tích, đánh giá, tìm hiểu ý nghĩa và trực quan hóa dữ liệu thông qua ngôn ngữ python và các thư viện hỗ trợ. Người chia sẻ nguồn dữ liệu này đồng ý với việc cho phép sử dụng nguồn dữ liệu cho mục đích thực hành và khám phá dữ liệu.

## Khám phá dữ liệu

* + Bộ dữ liệu được lưu dưới dạng một tập tin csv chứa một bảng thông tin duy nhất về các bệnh nhân được khảo sát.
  + Với mỗi dòng dữ liệu là thông tin về kết quả khảo sát của một bệnh nhân riêng biệt với 8762 bệnh nhân còn lại, dòng này cung cấp cho chúng ta giá trị tương ứng của 26 cột thuộc tính cho từng bệnh nhân. Tất cả các dòng đều dại diện cho cùng một loại thông tin, kết quả của từng dòng cũng được ghi đúng định dạng, kiểu dữ liệu. Trong trường hợp nếu có nhiều dòng mang ý nghĩa khác hoặc có cấu trúc khác biệt có thể mang lại khó khăn trong quá trình xử lý dữ liệu hoặc nghiêm trong hơn là làm sai kết các kết quả phân tích nếu không được phát hiện.
  + Bộ dữ liệu bao gồm một tập thuộc tính đa dạng về cả giá trị lẫn kiểu dữ liệu, cung cấp cho người sử dụng một nguồn thông tin đáng giá và cái nhìn tổng quan về vấn đề khi thực hiện phân tích. Các thuộc tính của dữ liệu và kiểu dữ liệu của chúng:
    - Patiend ID- chuỗi: mã định danh duy nhất cho mỗi bệnh nhân.
    - Age- số nguyên: tuổi của bệnh nhân.
    - Sex- chuỗi: Giới tính của bệnh nhân (Nam/Nữ).
    - Cholestoreol- số nguyên: Mức độ cholesterol trong người bệnh nhân.
    - Blood Pressure- chuỗi: Huyết áp tâm thu và huyết áp tâm trương của bệnh nhân.
    - Diabetes- logic: có bị tiểu đường hay không (1: Có, 0:không).
    - Family History-logic: người nhà có tiền sử bệnh tim hay không (1: Có, 0:không).
    - Smoking-logic: có thói quen hút thuốc hay không (1: Có, 0:không).
    - Obestity-logic: có bị béo phì hay không (1: Có, 0:không).
    - Alcohol Consumption- chuỗi: có tiêu thụ cồn hay không (1: Có, 0:không)
    - Exercie Hourse Per Week- số thực: số giờ luyện tập thể dục hằng tuần.
    - Diet-chuỗi: chế độ ăn uống(Khỏe mạnh-Thông thường-Không tốt).
    - Previous Heart Problems-logic: có từng có vấn đề về tim không (1: Có, 0:không).
    - Medication Use-logic: có sử dụng thuốc nói chung hay không (1: Có, 0:không).
    - Stress Level-số nguyên: mức độ stress được cho bởi bệnh nhân trên thang 1- 10.
    - Sedentary Hours Per Day-số thực: số giờ không vận động hằng ngày.
    - Income-Số thực: thu nhập của bệnh nhân.
    - BMI-số thực: chỉ số BMI của bệnh nhân.
    - Triglycerides-số nguyên: mức độ của chỉ số Triglycerides của bệnh nhân.
    - Physiscal Activity Days Per Week- Số nguyên: số ngày có hoạt động thể chất trong tuần.
    - Sleep Hours Per Day- số nguyên: số giờ ngủ mỗi ngày.
    - Country-chuỗi: quốc gia bệnh nhân sinh sống.
    - Continent-chuỗi: llục địa tương ứng với quốc gia của bệnh nhân.
    - Hemisphere-chuỗi: bán cầu của lục địa tương ứng.
    - Heart Attack Risk-logic: có mắc nhồi máu cơ tim không
  + Nhìn chung hầu như tất cả các thuộc tính đều có kiểu dữ liệu phù hợp với nhu cầu hiểu về dữ liệu với người sử dụng. Việc quy chuẩn một số thuộc tính như stress, Alcohol Consumption, diet vào các danh sách phân chia theo mức độ giúp dễ dàng hơn trong việc phân nhóm dữ liệu. Tuy nhiên điều này cũng có thể mang lại tác hại nếu cách phân loại của người tạo nên dữ liệu khác với người phân tích, nhiều khi ta không thể biết điều kiện, lí do và các nguyên nhân tạo

nên giá trị dữ liệu này.

* + Một số thuộc tính sẽ không được sử dụng vì không có tính khách quan như chỉ số BMI, Medication Use.
  + Nhóm chỉ thực hiện việc tiền xử lý dữ liệu đối với cột huyết áp của bệnh nhân. Hiện tại do cột dữ liệu này lưu giá trị huyết áp dưới dạng chuỗi “huyết áp tâm thu/huyết áp tâm trương”, việc này tuy đúng với các ghi chép thực tế nhưng không phù hợp với việc trực quan hay phân tích dữ liệu do đó sẽ cần được tách ra làm hai cột thuộc tính riêng là Systolic blood pressure và Diastolic blood pressure. Ta chỉ cần lấy ra các giá trị của cột Blood Pressure, thực hiện phân tích cú pháp chuỗi để lấy ra 2 danh sách giá trị huyết áp tương ứng và kết hợp lại với bảng dữ liệu ban đầu.

## Trực quan và ý nghĩa

## Trực quan dữ liệu

* + Bộ dữ liệu sau khi tiền xử lý tuy có rất nhiều quan sát và thuộc tính giúp mang lại rất nhiều trực quan có thể thực hiện tuy nhiên nhóm chỉ lựa chọn một số biểu đồ hỗ trợ trực quan các vấn đề đáng quan tâm. Các vấn đề nhóm cho rằng quan trọng:
    - **Các thói quen sinh hoạt, bệnh lý nào sẽ gây nên nguy cơ nhồi máu cơ tim ?** : Nhóm phân tích xem trong 5 yếu tố tiểu đường, tiền sử người thân, hút thuốc, béo phì, tiêu thụ rượu bia thì yếu tố nào là tác nhân quan trọng nhất. Biểu đồ Radar được sử dụng trong trường hợp này.
    - **Huyết áp của một cá nhân bị ảnh hưởng thế nào bởi thời gian họ ngủ ?** : Tình trạng thiếu ngủ do làm việc quá sức, thức đêm giải trí của giới trẻ dẫn đến việc thiếu ngủ có phải là nguyên nhân khiến gia tăng nguy cơ nhồi máu cơ tim hay không. Biểu đồ đường sẽ thể hiện mức huyết áp trung bình của các bệnh nhân có chung khoảng thời gian ngủ.
    - **Độ tuổi là nạn nhân phổ biến của nhồi màu cơ tim ?** : Việc biết được câu trả lời không chỉ giúp đánh giá tình hình mắc bệnh ở các độ tuổi mà đồng thời còn giúp cho người xem cân nhắc điều chỉnh thói quen sinh hoạt sao cho phù hợp với độ tuổi của mình dựa trên nguy cơ mắc ở độ tuổi này. Biểu đồ tròn sẽ giúp thể hiện tổng quan mức độ của từng độ tuổi so với tổng thể.
    - **Liệu việc tập thể dục thường xuyên có thực sự giảm nguy cơ nhồi máu cơ tim ?:** Nhóm muốn tìm hiểu xem liệu việc có thói quen luyện tập thể dục

đều đặn và chất lượng liệu có giúp con người tránh nguy cơ nhồi máu cơ tim hay không và liệu có áp dụng cho mọi độ tuổi. Biểu đồ heatmap sẽ cho cái nhìn tổng quan và so sánh cụ thể trong trường hợp này.

* + - **Liệu việc ngồi một chỗ trong thời gian dài có tăng nguy cơ nhồi máu cơ tim hay không ?:** Bên cạnh mục đích kiểm tra xem thói quen ít vận động của từng nhóm tuổi , ta đồng thời có thể xem liệu việc ngồi nhiều ít ra ngoài của con người có gây nên nguy cơ mắc bệnh không. Biểu đồ miền sẽ phù

hợp trong trường hợp này.

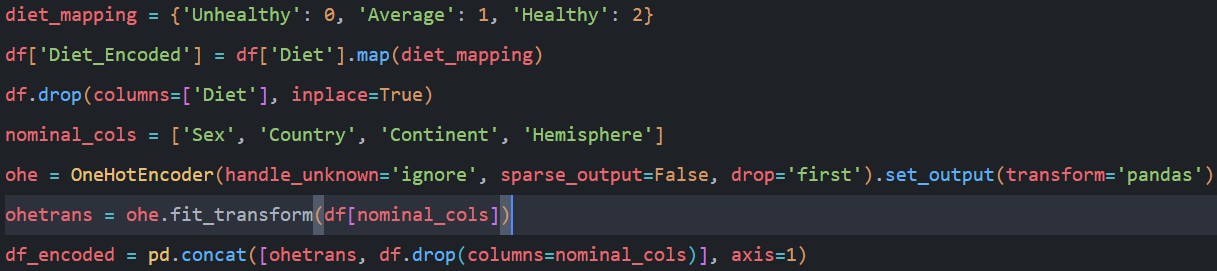
* + - **Mức độ stress của từng nhóm tuổi và ảnh hưởng đến nguy cơ nhồi màu cơ tim:** Stress là một từ khóa xuất hiện rất phổ biến trong nhiều năm trở lại đây, đặc biệt là ở người trẻ do nhiều vấn đề từ cuộc sống. Nhóm muốn kiểm chứng xem độ tuổi nào đang chịu ảnh hưởng nặng nề nhất từ stress và số lượng người mắc stress cao ( từ 6 trở lên). Và đặc biệt hơn là liệu đây có phải là một nhân tố quan trọng thầm lặng gây nên nhồi máu cơ tim. Biểu đồ heatmap sẽ cho cái nhìn toàn thể về hiện trạng này.
    - **Tiểu đường có đang trở thành căn bệnh chung của con người? :**Nhiều năm trở lại đây, do sự dày đặt của công việc và công nghệ, càng nhiều người trở nên mờ nhạt với các hoạt động thể chất từ trẻ nhỏ đến người già. Nhóm muốn kiểm tra xem ở độ tuổi nào thì bắt đầu có sự xuất hiện của bệnh béo phì. Biểu đồ tần suất (histogram) sẽ cho thấy sự phân bố của tập dữ liệu theo từng nhóm tuổi của các bệnh nhân mắc bệnh béo phì.
    - **Nhồi máu cơ tim có bị ảnh hưởng bởi vấn đề địa lý hay không và các bệnh nhân ở châu lục này sinh hoạt thế nào? :** Nhóm bên cạnh việc tham khảo xem môi trường sinh sống của các châu lục này có ảnh hưởng đến nhồi máu cơ tim hay không và thói quen sinh hoạt của những cá nhân ở vùng này như thế nào. **Biểu đồ Boxplot** được sử dụng trong trường hợp này.
    - **Mức độ triglycerides có tăng nguy cơ nhồi máu cơ tim hay không ? :** triglycerides là một chất dinh dưỡng cần thiết và phổ biến trong bữa ăn của con người tuy nhiên gần đây lại bị lạm dụng tiêu thụ quá mức làm tăng mức độ chất này có trong máu. Nhóm muốn kiểm tra xem tác động của chất này đến nguy cơ mắc nhồi màu cơ tim của hai giới tính nam và nữ. Biểu đồ Violin được sử dụng để có cái nhìn tổng quát về hai giới tính và hai nhóm có hay không mắc bệnh.
    - **Tiểu đường phổ biến như thế nào?** tương tự với chất béo, đường là một thực phẩm còn được sử dụng nhiều và rộng rãi hơn với mọi độ tuổi. Trước khi kiểm tra tiểu đường ảnh hưởng thế nào đến nhồi máu cơ tim, ta cần kiểm tra xem độ tuổi nào đang mắc bệnh tiểu đường nhiều nhất. Biểu đồ đường rất phù hợp cho việc quan sát sự thay đổi theo từng nhóm tuổi.
    - **Lối sống nhiều thói quen xấu ảnh hưởng thế nào đến nguy cơ mắc nhồi máu cơ tim? :** không chỉ kiểm thử tác động của từng thói quen đến nguy cơ mắc bệnh, ta cần xem thử việc liệu một người có nhiều hơn 1 thói quen xấu sẽ dẫn tới nguy cơ mắc bệnh như thế nào. Biểu đồ Veen rất phù hợp để xem giao thoa giữa các thói quen này, thể hiện bởi số lượng bệnh nhân mắc bệnh.
  + Chi tiết về các biểu đồ, loại biểu đồ, ý nghĩa, nguyên nhân sử dụng và kết luận của từng biểu đồ được mô tả tại các file ipynb kèm giải thích của mã nguồn.

## Mô hình dự đoán

* + **Bước 1:** Gọi các thư viện cần thiết cho xử lý, học máy bao gồm : scikit-learn,

pandas, numpy…

* + **Bước 2:** Mã hóa nhãn và mã hóa one hot. Bằng cách chuyển đổi cột thành giá trị số, mô hình sẽ tiếp nhận và xử lý dữ liệu dễ dàng hơn.



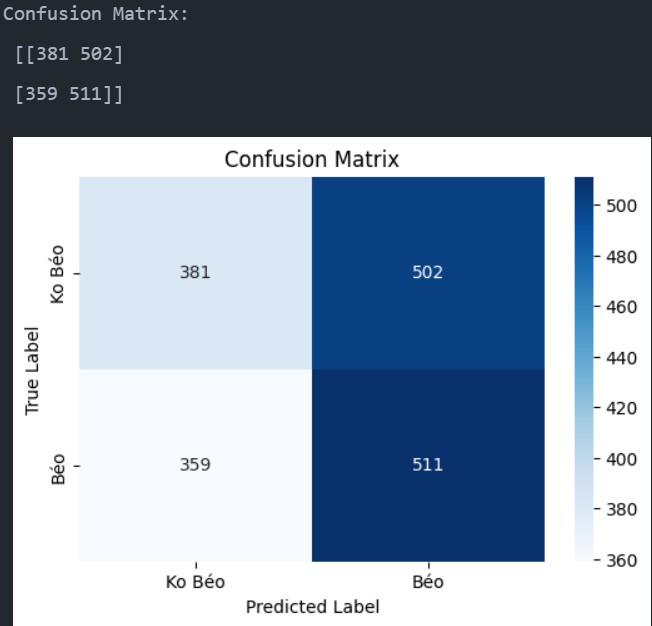
#### Bước 3:

* + - Tạo biến mục tiêu là tình trạng béo phì (Obesity).
    - Chia 20% dữ liệu để kiểm tra, 80% để huấn luyện.
    - Đảm bảo quá trình huấn luyện có thể tái lập.
    - Sử dụng mô hình huấn luyện hồi quy Logistic (LogisticRegression).
    - Sau cùng in độ chính xác của mô hình ra màn hình.



#### Bước 4:

* + - Hiện ma trận nhầm lẫn cho biết số lượng dự đoán đúng hoặc sai theo mô hình.
    - Kết quả dự đoán từ mô hình:
      * Độ chính xác (Accuracy): Đây là tỷ lệ các dự đoán đúng của mô hình so với tổng số dự đoán. Độ chính xác cao cho thấy mô hình phân loại tình trạng béo phì hiệu quả trong tập kiểm tra.
      * Ma trận nhầm lẫn (Confusion Matrix) cho biết chi tiết các loại dự đoán:
        + True Positives (TP): Số lượng trường hợp mà mô hình dự đoán đúng là "Béo" khi thực tế là "Béo".
        + True Negatives (TN): Số lượng trường hợp mà mô hình dự đoán đúng là "Không béo" khi thực tế là "Không béo".
        + False Positives (FP): Số lượng trường hợp mà mô hình dự đoán là "Béo" nhưng thực tế là "Không béo".
        + False Negatives (FN): Số lượng trường hợp mà mô hình dự đoán là "Không béo" nhưng thực tế là "Béo.



* + **Kết luận:** Mô hình hồi quy logistic này được thiết kế để ‘**dự đoán khả năng một người có bị béo phì hay không’** dựa trên các đặc trưng đầu vào từ dữ liệu. Cụ thể, mô hình sẽ dự đoán nhãn đầu ra là:
    - "Béo” nếu mô hình cho rằng khả năng người đó bị béo phì là cao.
    - "Không béo" nếu mô hình cho rằng người đó không bị béo phì.
  + **Ý nghĩa kết quả dự đoán:** Kết quả dự đoán từ mô hình giúp xác định khả năng béo phì của một người dựa trên các yếu tố đầu vào như chế độ ăn uống, giới tính, quốc gia, lục địa, và bán cầu sinh sống. Điều này có thể hỗ trợ trong các ứng dụng y tế và sức khỏe cộng đồng, cho phép xác định các nhóm đối tượng có nguy cơ cao và đưa ra các biện pháp phòng ngừa béo phì hiệu quả.

# PHẦN 4: TÀI LIỆU THAM KHẢO

## Tài liệu nhóm

1. Đường dẫn github: <https://github.com/Dzivilord/DV-TW-Lab01/tree/main>
2. Tài liệu thảo luận: [https://docs.google.com/document/d/1-](https://docs.google.com/document/d/1-9WZsXLLxO_FCZLZpiDrbraO5c1uuenu-wuuBlfSFsM/edit?hl=vi&tab=t.0) [9WZsXLLxO\_FCZLZpiDrbraO5c1uuenu-wuuBlfSFsM/edit?hl=vi&tab=t.0](https://docs.google.com/document/d/1-9WZsXLLxO_FCZLZpiDrbraO5c1uuenu-wuuBlfSFsM/edit?hl=vi&tab=t.0)

## Tài liệu tham khảo

1. Python Data Visualization Cookbook.
2. Aurélien Géron, ‘Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow’.
3. Kyle Gallatin & Chris Albon, ‘Machine Learning with Python Cookbook’.
4. ProgrammingKnowleage, ‘Matplotlib Tutorial for Begineers (Python)|Learn Data…’ [Trực tuyến]. Địa chỉ: https://[www.youtube.com/playlist?list=PLS1QulWo1RIZ3tcrdZodjuXTDTIIXH8](http://www.youtube.com/playlist?list=PLS1QulWo1RIZ3tcrdZodjuXTDTIIXH8) EW