

Giải thích chi tiết từng cột

1. Number of times pregnant (Số lần mang thai)

- **Giải thích:** Đây là số lần một người phụ nữ đã từng mang thai.
- **Tại sao lại quan trọng?** Mang thai là một gánh nặng sinh lý có thể ảnh hưởng đến khả năng dung nạp glucose của cơ thể. Phụ nữ mang thai nhiều lần có thể có nguy cơ cao hơn phát triển bệnh tiểu đường.

2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test (Nồng độ glucose huyết tương sau 2 giờ làm nghiệm pháp dung nạp glucose đường uống)

- **Giải thích:** Đây là chỉ số đường huyết được đo **2 giờ sau** khi bệnh nhân uống 75g dung dịch glucose. Đây là tiêu chuẩn vàng để chẩn đoán tiểu đường và tiền tiểu đường (như trong bảng 1 bạn đã xem).
- **Tại sao lại quan trọng?** Đây là thuộc tính **quan trọng nhất**. Chỉ số này trực tiếp phản ánh khả năng kiểm soát đường huyết của cơ thể. Giá trị càng cao, nguy cơ tiểu đường càng lớn.

3. Diastolic blood pressure (mm Hg) (Huyết áp tâm trương - đơn vị mmHg)

- **Giải thích:** Đây là áp lực trong động mạch khi tim giãn ra (là số dưới trong chỉ số huyết áp, ví dụ 120/80).
- **Tại sao lại quan trọng?** Huyết áp cao (cao huyết áp) và tiểu đường thường đi đôi với nhau như một phần của "Hội chứng chuyển hóa". Cao huyết áp là một yếu tố nguy cơ độc lập cho bệnh tiểu đường và các biến chứng của nó.

4. Triceps skin fold thickness (mm) (Độ dày nếp gấp da vùng cơ tam đầu - đơn vị mm)

- **Giải thích:** Đây là một phương pháp đo lượng mỡ dự trữ trong cơ thể. Người ta dùng một thước kẹp đặc biệt để kẹp và đo độ dày của một nếp da ở mặt sau cánh tay.
- **Tại sao lại quan trọng?** Chỉ số này đánh giá **tình trạng béo phì**, một yếu tố nguy cơ chính dẫn đến kháng insulin và tiểu đường type 2.

5. 2-Hour serum insulin (mu U/ml) (Nồng độ insulin huyết thanh sau 2 giờ - đơn vị micro đơn vị trên mililit)

- **Giải thích:** Đây là lượng insulin trong máu được đo cùng thời điểm với chỉ số đường huyết sau 2 giờ (ở cột 2).
- **Tại sao lại quan trọng?** Nó cho biết **mức độ phản ứng của cơ thể** với lượng glucose được nạp vào.
 - Nồng độ insulin cao + đường huyết cao: Cho thấy tình trạng **kháng insulin** (cơ thể sản xuất nhiều insulin nhưng không sử dụng hiệu quả).
 - Nồng độ insulin thấp + đường huyết cao: Cho thấy **suy giảm chức năng tế bào beta** của tuyến tụy (không sản xuất đủ insulin).

6. Body mass index (BMI) (Chỉ số khối cơ thể)

- **Giải thích:** $BMI = \text{Cân nặng (kg)} / (\text{Chiều cao (m)})^2$. Đây là chỉ số đánh giá thể trạng gầy/béo phổ biến.
- **Tại sao lại quan trọng?** Giống như cột 4, đây là một thước đo mạnh mẽ khác của **béo phì**, yếu tố nguy cơ trung tâm của tiểu đường type 2.

7. Diabetes pedigree function (Hàm phả hệ tiểu đường)

- **Giải thích:** Đây là một chỉ số **lượng hóa tiền sử gia đình** mắc bệnh tiểu đường. Nó được tính toán dựa trên mối quan hệ huyết thống và số lượng thành viên trong gia đình mắc bệnh. Giá trị càng cao cho thấy nguy cơ di truyền càng lớn.
- **Tại sao lại quan trọng?** Tiểu đường type 2 có yếu tố di truyền mạnh mẽ. Một người có tiền sử gia đình mắc bệnh sẽ có nguy cơ cao hơn.

8. Age (years) (Tuổi - đơn vị năm)

- **Giải thích:** Tuổi của bệnh nhân tại thời điểm khảo sát.
- **Tại sao lại quan trọng?** Nguy cơ mắc tiểu đường type 2 tăng lên theo tuổi tác, do sự gia tăng tình trạng kháng insulin và suy giảm chức năng tuyến tụy.

9. Class variable (0 or 1) (Biến mục tiêu - Nhân phân loại)

- **Giải thích:** Đây là **kết quả chẩn đoán** mà mô hình máy học cần dự đoán.
 - **Giá trị 0:** Âm tính với bệnh tiểu đường (Không mắc bệnh).
 - **Giá trị 1:** Dương tính với bệnh tiểu đường (Mắc bệnh).

- **Tại sao lại quan trọng?** Đây là "câu trả lời" cho mỗi bệnh nhân. Mô hình sẽ học mối quan hệ giữa 8 thuộc tính đầu tiên (đặc điểm) và kết quả này (nhãn) để sau đó dự đoán trên những bệnh nhân mới.

Tóm tắt ý nghĩa

8 thuộc tính đầu tiên là các **đặc điểm đầu vào (features)** phản ánh các yếu tố nguy cơ chính của bệnh tiểu đường type 2:

- **Rối loạn đường huyết** (trực tiếp): Cột 2.
- **Béo phì/Kháng insulin**: Cột 4, 5, 6.
- **Huyết áp** (yếu tố liên quan): Cột 3.
- **Di truyền**: Cột 7.
- **Nhân khẩu học** (Tuổi, Tiền sử sinh sản): Cột 1, 8.

Thuộc tính thứ 9 là **mục tiêu cần dự đoán (target)**

Các tên biến sử dụng khi code với dataset

Cột	Tên cột
1	pregnancies
2	glucose
3	blood_pressure
4	skin_thickness
5	insulin
6	bmi
7	dpf
8	age
9	class