Data analysis:

Thông tin dữ liệu.

Bộ dữ liệu "Default of Credit Card Clients" (ID: 350 trên UCI Machine Learning Repository) tập trung vào việc dự đoán rủi ro tín dụng của khách hàng thẻ tín dụng tại Đài Loan. Cụ thể, mục tiêu là dự đoán liệu khách hàng có khả năng không thanh toán khoản nợ trong tháng tới hay không, dựa trên các đặc điểm nhân khẩu học, tài chính, và lịch sử tín dụng của họ.

Chi tiết dữ liệu:

Nguồn: Bộ dữ liệu này được cung cấp bởi một nghiên cứu của các nhà khoa học Đài Loan và được công bố trong một bài báo học thuật.

Mục đích: Hỗ trợ các ngân hàng, tổ chức tài chính trong việc đánh giá rủi ro tín dụng và quản lý danh mục đầu tư.

Biến số (variables):

Đầu vào (features): Bao gồm 23 biến, chẳng hạn:

Thông tin cá nhân: Tuổi, giới tính, tình trạng hôn nhân, trình độ học vấn.

Tình trạng tín dụng: Số dư tín dụng, lịch sử thanh toán qua các tháng.

Mức tiêu dùng/thanh toán: Tổng tiền chi tiêu/thanh toán hàng tháng.

Đầu ra (target): Một biến nhị phân (0 hoặc 1) cho biết khách hàng có mặc định (default) trong việc trả nợ tháng tới hay không.

Mô tả về nguồn dữ liệu: Default of Credit Card Clients Dataset

1. Giới thiệu chung

Tên tập dữ liệu: Default of Credit Card Clients Dataset

Mã UCI: 350

Liên kết UCI Repository: UCI Dataset

Liên kết tải dữ liệu: Data CSV

Mục tiêu: Nghiên cứu khả năng dự đoán việc khách hàng có mặc định thanh toán thẻ tín dụng hay không, dựa trên các biến giải thích như lịch sử thanh toán, giới tính, trình độ học vấn, tình trạng hôn nhân, và các thông tin tài chính khác.

2. Mô tả dữ liệu

Số lượng mẫu: 30,000

Số lượng đặc trưng (features): 23

Loại bài toán: Phân loại (Classification)

Đặc trưng dữ liệu: Dữ liệu đa biến (Multivariate), dạng số nguyên (Integer) và số thực (Real).

Không có giá trị thiếu.

3. Mục tiêu và ý nghĩa

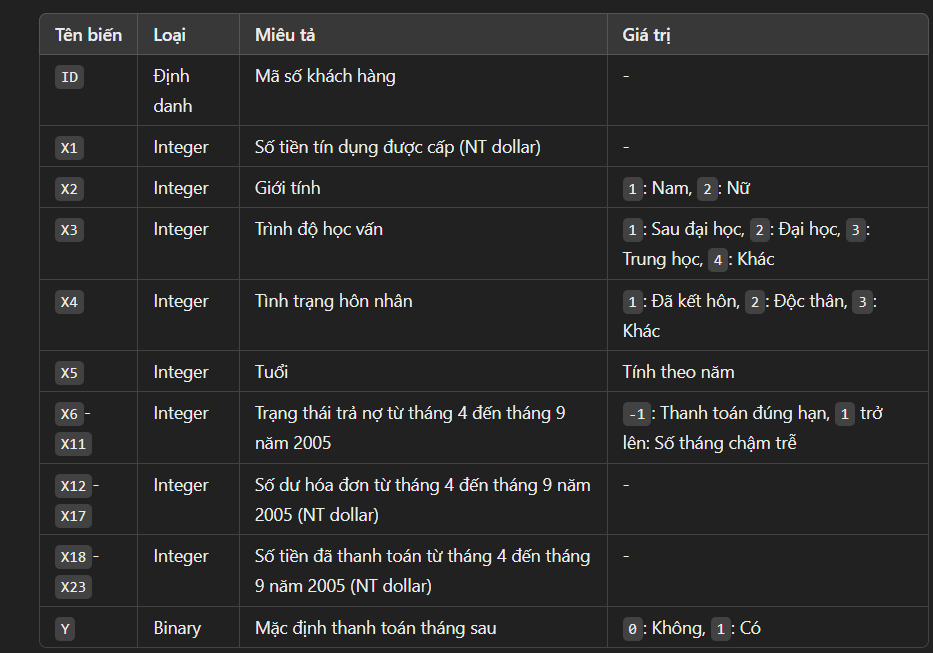
Biến mục tiêu (Y):

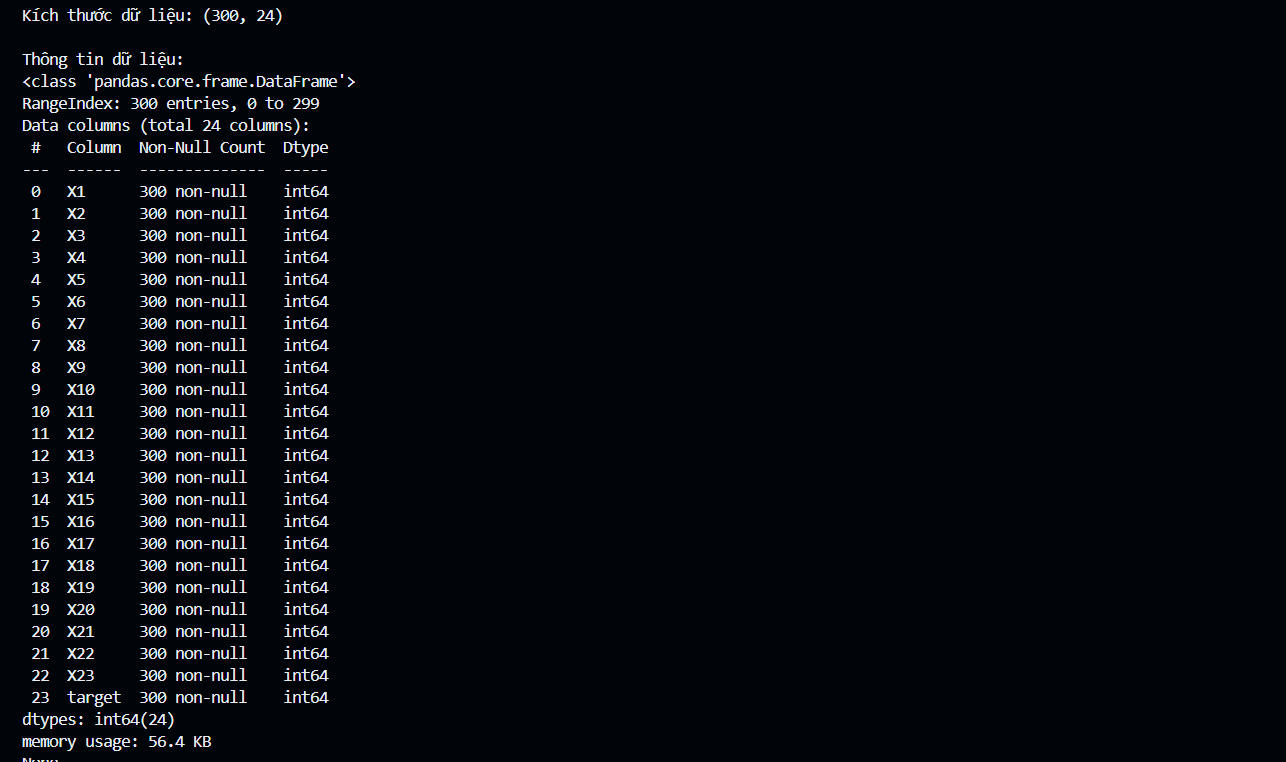
Là biến nhị phân, đại diện cho việc khách hàng có mặc định thanh toán vào tháng tiếp theo hay không:

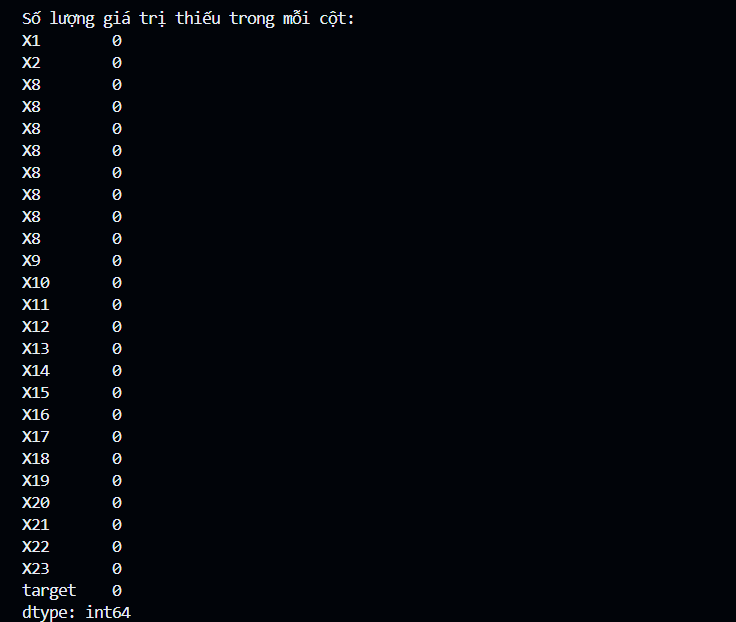
1: Có mặc định (Default).

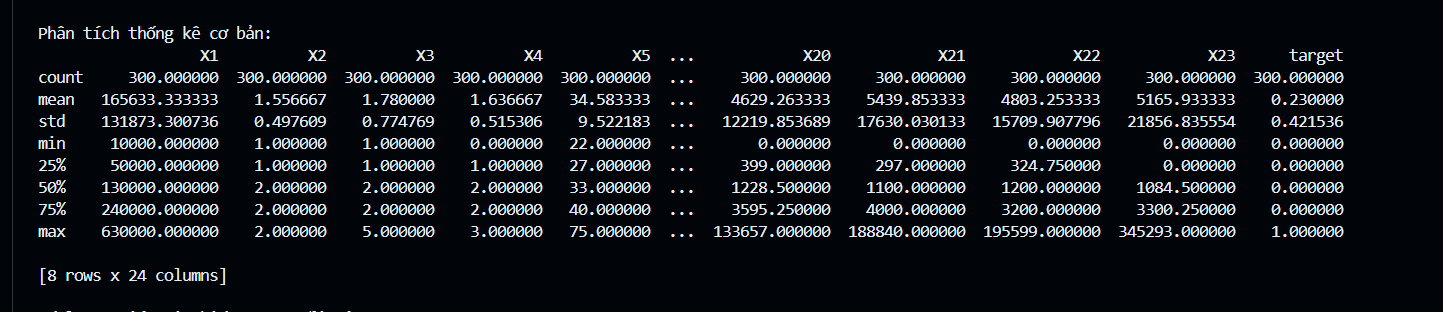
0: Không mặc định (Non-default).

Ý nghĩa thực tiễn: Từ góc độ quản lý rủi ro, dự đoán chính xác khả năng mặc định thanh toán giúp các tổ chức tài chính tối ưu hóa quản lý tín dụng và giảm rủi ro.









**Phân tích thống kê mô tả :**

Các biến nhân khẩu học (Demographics):

Giới tính (X2): Phân phối giữa nam và nữ.

Trình độ học vấn (X3): Tỷ lệ các cấp độ học vấn (sau đại học, đại học, trung học, khác).

Tình trạng hôn nhân (X4): Phân bố các trạng thái (đã kết hôn, độc thân, khác).

Tuổi (X5): Phân phối tuổi khách hàng, các nhóm tuổi (ví dụ: <25, 25–40, >40).

Hình Ảnh

