

# BTLT- Day5 - Credit Scoring System

TranHuy

## Bối cảnh đề bài

Trong bối cảnh kinh tế có nhiều biến động, việc đánh giá chính xác khả năng trả nợ của khách hàng trở thành yếu tố sống còn đối với các tổ chức tín dụng. Một ngân hàng thương mại đang phát triển hệ thống credit scoring tự động để hỗ trợ quyết định cho vay. Ngân hàng đã thu thập được dataset gồm 20,000 hồ sơ khách hàng từ 3 năm gần đây, với biến mục tiêu Default (0: trả nợ đầy đủ, 1: vỡ nợ/chậm trả > 90 ngày). Tỷ lệ vỡ nợ thực tế là 12%.

## Các biến đầu vào

- **Income:** Thu nhập hàng tháng (triệu VND)
- **Age:** Tuổi của khách hàng
- **CreditScore:** Điểm tín dụng từ CIC (300-850)
- **DebtRatio:** Tỷ lệ nợ trên thu nhập (%)

## Kết quả mô hình Logistic Regression

Sau khi huấn luyện:

$$\begin{aligned} \log\text{-odds} = & 2.8 - 0.0003 \times \text{Income} - 0.025 \times \text{Age} \\ & - 0.006 \times \text{CreditScore} + 0.018 \times \text{DebtRatio} \end{aligned} \quad (1)$$

Các hệ số:

$$\begin{aligned} \beta_0 \text{ (Intercept)} &= 2.8 \\ \beta_1 \text{ (Income)} &= -0.0003 \\ \beta_2 \text{ (Age)} &= -0.025 \\ \beta_3 \text{ (CreditScore)} &= -0.006 \\ \beta_4 \text{ (DebtRatio)} &= 0.018 \end{aligned}$$

## Hiệu suất mô hình

- $\text{AUC} = 0.78$
- $\text{Precision} = 0.55$
- $\text{Recall} = 0.85$

## Chính sách chi phí của ngân hàng

- Chi phí opportunity cost khi từ chối nhằm khách hàng tốt = 3 triệu VND
- Chi phí tổn thất khi cho vay nhằm khách hàng xấu = 50 triệu VND

## Câu hỏi

1.

Biết khách hàng A 45 tuổi có thu nhập 25.000.000 VND/tháng, điểm tín dụng 680 và tỉ lệ nợ/thu nhập là 30%. Nếu ngân hàng đặt ngưỡng từ chối là 13% thì khách hàng A có được chấp nhận hay không? Vì sao?

2.

Với thông tin về Precision = 0.55, Recall = 0.85 và chính sách chi phí của ngân hàng, hãy:

- a) Tính toán confusion matrix cho toàn bộ 20,000 khách hàng
- b) Tính expected cost tổng thể và chi phí trung bình trên mỗi khách hàng

### **Yêu cầu**

Trình bày ngắn gọn các bước của bài làm.

#### **0.1 Công thức:**

$$\text{True Positive} = \text{Recall} \times \text{Actual Positive} \quad (1)$$

$$\text{False Positive} = \frac{\text{True Positive}}{\text{Precision}} - \text{True Positive} \quad (2)$$

$$\text{False Negative} = \text{Actual Positive} - \text{True Positive} \quad (3)$$

$$\text{True Negative} = \text{Actual Negative} - \text{False Positive} \quad (4)$$

$$\text{Expected Cost} = (\text{False Negative} \times \text{Chi phí bỏ sót khách hàng xấu}) \quad (5)$$

$$+ (\text{False Positive} \times \text{Chi phí từ chối nhầm khách hàng tốt}) \quad (6)$$

Hình 1: