

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

BÁO CÁO ĐỒ ÁN

**MÔN HỌC: LẬP TRÌNH PYTHON CHO MÁY HỌC**

**ĐĂNG KÝ ĐỀ TÀI**

**Giảng viên hướng dẫn:** TS. Nguyễn Vinh Tiệp

**Sinh viên thực hiện:** Huỳnh Việt Tiến - 23521570

Nguyễn Công Phát - 23521143

Phạm Phú Bảo - 23520133

Nguyễn Minh Bảo - 23520123

**Lớp học phần:** CS116.P22

**Thành phố Hồ Chí Minh, ngày 30 tháng 3 năm 2025**

**Mục lục**

**I/ Danh sách thành viên 3**

**II/ Giới thiệu đề tài 3**

# I/ Danh sách thành viên

* + - Huỳnh Việt Tiến - MSSV: 23521570
    - Nguyễn Công Phát - MSSV: 23521143
    - Phạm Phú Bảo - MSSV: 23520133
    - Nguyễn Minh Bảo - MSSV: 23520123

# II/ Giới thiệu đề tài:

1. **Tên bộ dữ liệu:**

- ***Dự đoán phê duyệt khoản vay***

- Bộ dữ liệu này là một tập hợp thông tin liên quan đến các ứng viên vay vốn, được sử dụng để dự đoán khả năng phê duyệt khoản vay. Dữ liệu bao gồm các đặc điểm cá nhân và tài chính của ứng viên, như tuổi, thu nhập, tình trạng sở hữu nhà, thời gian làm việc, mục đích vay vốn, loại khoản vay, số tiền vay, lãi suất khoản vay, tỷ lệ phần trăm thu nhập dành cho khoản vay, lịch sử tín dụng và trạng thái phê duyệt khoản vay (loan\_status).

1. **Lý do chọn bộ dữ liệu:**

- Đề tài "Dự đoán phê duyệt khoản vay" là một lựa chọn cực kỳ thú vị và hữu ích để học tập và rèn luyện các kỹ năng liên quan đến bộ môn trí tuệ nhân tạo. Đây là một bài toán thực tế trong lĩnh vực tài chính, nơi mà việc dự đoán khả năng phê duyệt khoản vay dựa trên thông tin khách hàng không chỉ giúp ngân hàng tối ưu hóa quy trình mà còn giảm thiểu rủi ro tín dụng. Thông qua đề tài này, nhóm chúng em có cơ hội làm việc với dữ liệu đa dạng, từ dữ liệu số đến dữ liệu phân loại, và áp dụng các thuật toán học máy để giải quyết bài toán. Điều này không chỉ giúp nhóm chúng em hiểu rõ hơn về cách các mô hình hoạt động mà còn cải thiện kỹ năng xử lý dữ liệu, phân tích và đánh giá mô hình. Hơn nữa, đây là một bài toán mang tính ứng dụng cao, nên nhóm chúng em cảm thấy rất hào hứng khi biết rằng những gì mình học được có thể áp dụng vào thực tế sau này.

1. **Loại bài toán và các thông tin về bộ dữ liệu:**

- Bài toán này là bài toán ***phân loại 2 nhãn***, nhiệm vụ của chúng em sẽ phân loại xem người vay có đủ tiêu chuẩn để có thể được vay hay không dựa vào 12 đặc trưng có trong bộ dữ liệu. Người vay được chấp nhận vay là “1”, ngược lại nhãn không được chấp nhận vay là “0”.

- Bộ dữ liệu này bao gồm thông tin chi tiết về các ứng viên vay vốn, với tổng cộng **2,113 hàng** và **13 (cột)**. **Số đặc trưng có trong dataset là 12**. Các đặc trưng được chia thành hai loại chính là đặc trưng số và đặc trưng phân loại và một đặc trưng mục tiêu:

**3.1) Đặc trưng số (số lượng 6):**

* person\_age: Tuổi của ứng viên.
* person\_income: Thu nhập của ứng viên.
* loan\_amnt: Số tiền vay.
* loan\_int\_rate: Lãi suất khoản vay.
* loan\_percent\_income: Tỷ lệ phần trăm thu nhập dành cho khoản vay.
* cb\_person\_cred\_hist\_length: Thời gian lịch sử tín dụng của ứng viên.

**3.2) Đặc trưng phân loại(số lượng 5):**

* person\_home\_ownership: Tình trạng sở hữu nhà (RENT, OWN, MORTGAGE, OTHER).
* person\_emp\_length: Thời gian làm việc (năm).
* loan\_intent: Mục đích vay (EDUCATION, MEDICAL, VENTURE, PERSONAL, DEBTCONSOLIDATION).
* loan\_grade: Xếp hạng khoản vay (A, B, C, D, E).
* cb\_person\_default\_on\_file: Lịch sử vỡ nợ tín dụng (Y/N).

**3.3) Đặc trưng mục tiêu:**

* loan\_status: Trạng thái phê duyệt khoản vay (0: không được duyệt, 1: được duyệt).