

BÁO CÁO PHÂN TÍCH NGHIỆP VỤ (BUSINESS ANALYTICS)

Đề tài: Ứng dụng quy trình CRISP-DM trong bài toán tự động hóa phê duyệt khoản vay ngân hàng

Mục tiêu của báo cáo này là trình bày cách áp dụng quy trình CRISP-DM để giải quyết bài toán phân loại khách hàng vay vốn thành hai nhóm: "An toàn" và "Rủi ro", từ đó hỗ trợ ngân hàng giảm thiểu nợ xấu và nâng cao hiệu quả ra quyết định.

Thông tin sinh viên

Họ và tên	MSSV
Dương Quốc Vinh	523H0195
Phạm Nguyễn Thanh Uyên	521H0329
Lưu Chí Bảo	524H0144
Nguyễn Minh Triết	523H0107

Phần 1: Áp dụng quy trình CRISP-DM

1. Business Understanding

Mục tiêu kinh doanh của ngân hàng là giảm tỷ lệ nợ xấu và rủi ro tín dụng trong quá trình cho vay. Việc phê duyệt khoản vay thủ công thường tốn nhiều thời gian và phụ thuộc vào kinh nghiệm cá nhân, dễ dẫn đến sai sót.

Về mục tiêu khai phá dữ liệu, bài toán hướng tới việc xây dựng một mô hình có khả năng dự đoán chính xác khả năng rủi ro của khách hàng dựa trên dữ liệu lịch sử. Kết quả mô hình sẽ hỗ trợ cán bộ tín dụng ra quyết định nhanh và nhất quán hơn.

2. Data Understanding

Dữ liệu đóng vai trò nền tảng trong toàn bộ dự án. Ngân hàng cần thu thập và hiểu rõ các thuộc tính liên quan đến hành vi vay vốn của khách hàng, bao gồm:

- Thu nhập hàng tháng
- Nghề nghiệp và thời gian công tác

- Lịch sử tín dụng (nợ quá hạn, số lần vay trước đó)
- Tuổi, tình trạng hôn nhân
- Số khoản vay đang tồn tại

Việc hiểu rõ ý nghĩa và mối quan hệ giữa các thuộc tính giúp xác định dữ liệu nào thực sự có giá trị cho bài toán phân loại rủi ro.

3. Data Preparation

Đây là giai đoạn chiếm nhiều thời gian nhất trong dự án (khoảng 85%). Các công việc chính bao gồm làm sạch dữ liệu, xử lý giá trị bị thiếu, chuẩn hóa dữ liệu và loại bỏ dữ liệu nhiễu.

Ví dụ, các giá trị thu nhập bị thiếu có thể được thay thế bằng giá trị trung bình hoặc trung vị. Các bản ghi không hợp lệ hoặc trùng lặp cần được loại bỏ. Việc chuẩn bị dữ liệu tốt giúp mô hình học chính xác và giảm sai lệch kết quả.

4. Model Building

Trong bài toán này, phương pháp Decision Trees được đề xuất sử dụng. Decision Trees có ưu điểm là dễ hiểu, dễ giải thích và phổ biến trong lĩnh vực tài chính.

Mô hình cho phép biểu diễn các quy tắc ra quyết định dưới dạng cây, giúp ngân hàng dễ dàng giải thích lý do tại sao một khách hàng được xếp vào nhóm "An toàn" hoặc "Rủi ro", từ đó tăng tính minh bạch trong phê duyệt khoản vay.

5. Testing and Evaluation

Để đánh giá mô hình, có thể sử dụng các tiêu chí như độ chính xác (Accuracy), độ phủ (Recall) và độ chính xác dự đoán dương (Precision).

Trong bối cảnh ngân hàng, Recall đối với nhóm "Rủi ro" thường quan trọng hơn, vì việc bỏ sót khách hàng rủi ro có thể gây thiệt hại tài chính lớn. Do đó, mô hình cần được đánh giá dựa trên nhiều tiêu chí thay vì chỉ một chỉ số.

6. Deployment

Sau khi mô hình đạt yêu cầu, nó sẽ được tích hợp vào hệ thống công nghệ thông tin của ngân hàng. Khi khách hàng nộp hồ sơ vay, hệ thống sẽ tự động thu thập dữ liệu đầu vào và đưa ra kết quả phân loại rủi ro.

Cán bộ tín dụng có thể sử dụng kết quả này như một công cụ hỗ trợ ra quyết định, giúp rút ngắn thời gian xử lý hồ sơ và nâng cao hiệu quả quản lý rủi ro.