

TRƯỜNG ĐẠI HỌC SÀI GÒN
KHOA CÔNG NGHỆ THÔNG TIN



MÔN: PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC
BÀI 4: KHẢO SÁT ĐỀ TÀI

GV: Đỗ Như Tài

Nhóm lớp học: Nhóm 02 (Thứ 7 tiết 4+5)

Sinh viên thực hiện:

Họ tên	Mã sinh viên
Nguyễn Tuyết Nhi	3122410282
Trương Thành Lâm	3122410209
H' Như Lưk	3122410226

Thành phố Hồ Chí Minh, tháng 3 năm 2025

TÊN ĐỀ TÀI: PHÂN TÍCH VÀ ĐÁNH GIÁ RỦI RO TÍN DỤNG VỚI GERMAN CREDIT DATASET

GIỚI THIỆU

CÂU HỎI GỢI Ý:

- Lĩnh vực và nghiên cứu liên quan đã và đang phát triển như thế nào?
- Các vấn đề và bài toán đặt ra cần giải quyết là gì?

Ngày nay, rủi ro tín dụng là một vấn đề quan trọng trong lĩnh vực tài chính và ngân hàng. Với sự phát triển nhanh chóng của công nghệ dữ liệu và trí tuệ nhân tạo, việc phân tích và đánh giá rủi ro tín dụng ngày càng trở nên hiệu quả hơn. Thật vậy, trong bối cảnh nền kinh tế hiện đại, các tổ chức tài chính phải đối mặt với nhiều thách thức trong việc xác định khả năng hoàn trả khoản vay của khách hàng. Điều này đặt ra yêu cầu cấp thiết trong việc ứng dụng các phương pháp khai thác dữ liệu để phân tích rủi ro tín dụng.

Việc sử dụng các kỹ thuật khai thác dữ liệu và học máy giúp tối ưu hóa quy trình đánh giá rủi ro, hỗ trợ ngân hàng và tổ chức tài chính trong việc ra quyết định cho vay một cách chính xác hơn. Các mô hình phân tích rủi ro tín dụng giúp xác định các yếu tố quan trọng ảnh hưởng đến khả năng trả nợ của khách hàng, từ đó đề xuất các biện pháp quản lý rủi ro hiệu quả hơn. Một trong những bộ dữ liệu phổ biến được sử dụng để nghiên cứu vấn đề này là German Credit Dataset.

German Credit Dataset là một tập dữ liệu chuẩn được sử dụng rộng rãi trong nghiên cứu về rủi ro tín dụng. Tập dữ liệu này chứa thông tin về 1000 khách hàng với các đặc điểm khác nhau như tuổi tác, nghề nghiệp, tình trạng hôn nhân, số tiền vay, và lịch sử tín dụng. Mục tiêu của nghiên cứu là phân tích dữ liệu này để xác định các yếu tố chính ảnh hưởng đến quyết định tín dụng và đánh giá mức độ rủi ro của từng khách hàng.

Tại Việt Nam, vấn đề rủi ro tín dụng đang được quan tâm đặc biệt trong ngành ngân hàng. Các tổ chức tài chính đã bắt đầu áp dụng các phương pháp học máy và khai thác dữ liệu để cải thiện độ chính xác trong việc đánh giá tín dụng. Tuy nhiên, vẫn còn nhiều thách thức liên quan đến tính minh bạch của mô hình, độ chính xác của dữ liệu, và tính phù hợp với bối cảnh kinh tế trong nước.

Do đó, bài toán đặt ra là có thể áp dụng các kỹ thuật khai thác dữ liệu trên German Credit Dataset để xây dựng một mô hình phân tích và đánh giá rủi ro tín dụng phù hợp với bối cảnh thực tế của Việt Nam hay không?

MỤC TIÊU VÀ PHẠM VI NGHIÊN CỨU

Câu hỏi gợi ý:

- Mục tiêu nghiên cứu chính của đề tài là gì?
- Phạm vi nghiên cứu là gì?

Hướng dẫn:

- Đặt bài toán giải quyết và trình bày mục tiêu nghiên cứu chính của đề tài.
- Nếu phạm vi nghiên cứu của đề tài, bao gồm việc giới hạn phạm vi nghiên cứu và triển khai, các giả định ban đầu đối với nghiên cứu.

Với bài toán, vấn đề nêu ra ở trên thì mục tiêu của đề tài chúng tôi là phân tích và đánh giá rủi ro tín dụng bằng cách áp dụng các kỹ thuật khai thác dữ liệu vào German Credit Dataset nhằm hỗ trợ các tổ chức tài chính trong việc quản lý rủi ro tín dụng một cách hiệu quả hơn. Cụ thể, đề tài sẽ tập trung vào các mục tiêu chính sau:

- **Xác định các yếu tố quan trọng ảnh hưởng đến quyết định tín dụng** thông qua phân tích dữ liệu;
- **Ứng dụng các mô hình phân loại và dự đoán rủi ro tín dụng** để đánh giá khả năng vỡ nợ của khách hàng;
- **ok** và ứng dụng thực tế trong hệ thống tài chính Việt Nam.

Về phạm vi nghiên cứu: Đề tài sẽ tập trung vào việc phân tích dữ liệu từ German Credit Dataset và áp dụng các phương pháp khai thác dữ liệu để xây dựng mô hình đánh giá rủi ro tín dụng. Nghiên cứu được thực hiện trong phạm vi học thuật tại bộ môn Khoa học Dữ liệu, Khoa Công nghệ Thông tin – Trường Đại học Sài Gòn.

CÁC GIẢ THUYẾT VÀ CÁCH TIẾP CẬN

Giả thuyết nghiên cứu

Có thể xây dựng một mô hình phân tích và đánh giá rủi ro tín dụng dựa trên tập dữ liệu German Credit, thỏa mãn các yêu cầu sau:

- Ứng dụng các thuật toán khai thác dữ liệu để phân loại khách hàng có nguy cơ vỡ nợ cao và thấp.

- Nhận diện các đặc điểm quan trọng (thu nhập, lịch sử tín dụng, nghề nghiệp,...) ảnh hưởng đến khả năng hoàn trả khoản vay.
- Hỗ trợ ngân hàng và tổ chức tín dụng đưa ra quyết định cấp tín dụng một cách hiệu quả hơn.

Bài toán thứ nhất:

Phân tích các yếu tố ảnh hưởng đến rủi ro tín dụng bằng các phương pháp thống kê và trực quan hóa dữ liệu.

Bài toán thứ hai:

Ứng dụng các kỹ thuật phân cụm để nhóm khách hàng theo mức độ rủi ro, từ đó đề xuất chiến lược quản lý tín dụng hiệu quả.

- Với bài toán thứ nhất, chúng tôi tiến hành tiền xử lý dữ liệu, trích xuất thông tin từ các biến, sau đó áp dụng phân tích thống kê để xác định các yếu tố quan trọng ảnh hưởng đến khả năng hoàn trả khoản vay.
- Với bài toán thứ hai, chúng tôi thử nghiệm các thuật toán phân cụm như K-means, DBSCAN và Hierarchical Clustering để phân nhóm khách hàng dựa trên đặc điểm tài chính và lịch sử tín dụng.

Cách tiếp cận

Để kiểm chứng các giả thuyết trên, nghiên cứu áp dụng các phương pháp sau:

- **Tiền xử lý dữ liệu:** Làm sạch và chuẩn hóa dữ liệu từ German Credit Dataset để đảm bảo tính nhất quán.
- **Phân tích thống kê:** Sử dụng các phương pháp thống kê để xác định mối quan hệ giữa các biến và rủi ro tín dụng.
- **Kỹ thuật phân cụm:** Áp dụng các thuật toán phân cụm như K-means, DBSCAN hoặc Hierarchical Clustering để phân loại khách hàng theo mức độ rủi ro.
- **Đánh giá mô hình:** So sánh hiệu suất của các phương pháp để xác định mô hình phù hợp nhất với dữ liệu.

TÀI LIỆU THAM KHẢO

- Nguyễn, V. A., & Trần, B. T. (2018). *Phân cụm dữ liệu và ứng dụng trong phân tích rủi ro tín dụng*. Nhà xuất bản Khoa học và Kỹ thuật.
- Phạm, H. N. (2020). *Ứng dụng K-means trong phân nhóm khách hàng vay tín dụng*. Tạp chí Công nghệ Thông tin và Truyền thông, 15(4), 42–51.
- Trần, D. Q. (2019). *Phân tích dữ liệu lớn trong ngành tài chính - ngân hàng*. Nhà xuất bản Thống kê.