

Đề bài chủ đề Finance

Loan Default Prediction Analysis – Phân tích Rủi ro Khoản vay

Thử thách dữ liệu thực chiến thuộc cuộc thi **NextGen Analytics Challenge 2025**, dành cho bất kỳ ai yêu thích và đam mê lĩnh vực dữ liệu – đặc biệt là ứng dụng trong ngành Tài chính – Ngân hàng.

A. Nội dung đề bài

Chủ đề Finance

Loan Default Prediction Analysis – Phân tích Rủi ro Khoản vay

Bối cảnh bài toán, vai trò và mục tiêu phân tích

Bối cảnh

Bạn đang làm việc tại **phòng Phân tích Rủi ro Tín dụng** của một **ngân hàng thương mại lớn**, nơi cung cấp các sản phẩm vay tiêu dùng và vay kinh doanh cho khách hàng cá nhân.

Ban Giám đốc vừa yêu cầu bộ phận bạn thực hiện một báo cáo **phân tích tổng thể về đặc điểm hồ sơ khách hàng và các yếu tố rủi ro ảnh hưởng đến khả năng vỡ nợ khoản vay**. Mục đích là:

- Đánh giá các xu hướng rủi ro hiện hữu trong danh mục cho vay.
- Xác định nhóm khách hàng cần ưu tiên kiểm soát rủi ro.
- Đưa ra thông tin nền tảng cho việc xây dựng chính sách xét duyệt và lãi suất phù hợp.

Vai trò của bạn

Bạn đóng vai trò:

Chuyên viên phân tích dữ liệu tín dụng (Credit Risk Data Analyst).

Bạn chịu trách nhiệm:

- Làm sạch dữ liệu, trực quan hóa và phân tích các biến liên quan đến khả năng vỡ nợ.
- Xây dựng dashboard Power BI trực quan, dễ sử dụng cho lãnh đạo và phòng nghiệp vụ.

Người xem báo cáo:

- Ban Giám đốc (CEO, Giám đốc Tài chính, Giám đốc Rủi ro).
- Phòng Tín dụng Cá nhân (nhân viên phê duyệt khoản vay, quản lý danh mục).
- Phòng Phân tích và Kiểm soát Rủi ro.

Mục đích báo cáo

Báo cáo và dashboard Power BI cần trả lời **các câu hỏi cụ thể và nhất quán**, phục vụ mục tiêu:

- **Hiểu rõ đặc điểm hồ sơ khách hàng:**
 - Nhân khẩu học (tuổi, thu nhập, tình trạng hôn nhân, giáo dục).
 - Tình trạng công việc, mục đích vay, các yếu tố liên quan.
- **Đo lường và trực quan hóa tỷ lệ vỡ nợ:**
 - Tỷ lệ default chung toàn bộ danh mục.
 - So sánh tỷ lệ default giữa các nhóm khách hàng.
- **Xác định yếu tố rủi ro chính:**
 - Biến nào liên quan mạnh nhất đến khả năng vỡ nợ?
 - Nhóm khách hàng nào có rủi ro cao hơn mức trung bình?
- **Hỗ trợ ra quyết định:**
 - Đề xuất nhóm khách hàng nên ưu tiên kiểm soát rủi ro (theo Credit Score, DTI Ratio...).
 - Đề xuất chính sách xét duyệt/ giám sát.

Các yếu tố cần đạt được sau khi phân tích

Để đảm bảo tất cả thí sinh dự thi cùng trả lời một vấn đề thống nhất, báo cáo bắt buộc phải bao gồm:

- **Phân tích tổng quan danh mục khoản vay:**
 - Số lượng khoản vay, tỷ lệ vỡ nợ trung bình.
 - Phân bố khách hàng theo các biến chính (tuổi, thu nhập, điểm tín dụng, mục đích vay).
- **So sánh tỷ lệ vỡ nợ giữa các nhóm đặc trưng:**
 - Ví dụ: so sánh theo tình trạng hôn nhân, nghề nghiệp, mục đích vay, có đồng bảo lãnh không.
 - Trực quan hóa qua biểu đồ cột/ so sánh.
- **Trực quan hóa mối liên hệ giữa biến số và default:**
 - Ví dụ: mối tương quan giữa Credit Score và tỷ lệ default.
 - Phân tích Debt-to-Income Ratio với khả năng vỡ nợ.
- **Bảng tổng hợp các yếu tố rủi ro chính:**
 - Tìm ra ít nhất **3 yếu tố** có liên quan mạnh nhất đến khả năng vỡ nợ.
- **Đề xuất hướng hành động:**
 - Nêu ít nhất **2 khuyến nghị nghiệp vụ** dựa trên kết quả phân tích (ví dụ: tăng giám sát nhóm khách hàng nào, điều chỉnh điều kiện vay...).
- **Dashboard Power BI:**
 - Có tối thiểu:
 - Trang tổng quan toàn bộ danh mục.
 - Trang chi tiết tỷ lệ default theo nhóm biến chính.
 - Bộ lọc (slicer) để lựa chọn phân khúc khách hàng.
 - Tooltips hoặc highlight dữ liệu.
 - KPI Cards thể hiện số lượng khoản vay và tỷ lệ default tổng.

Mô tả Dataset

Dịch vụ cho vay tài chính được nhiều tổ chức trong các ngành khác nhau sử dụng, từ các ngân hàng lớn, tổ chức tín dụng đến các quỹ cho vay của chính phủ.

Một trong những mục tiêu chính của các tổ chức này là giảm thiểu tỷ lệ **khoản vay không thanh toán đúng hạn**, đồng thời đảm bảo khách hàng có khả năng hoàn trả khoản vay đúng cam kết. Để thực hiện điều này một cách hiệu quả và hệ thống, họ cần phân tích sâu dữ liệu lịch sử, xác định các đặc điểm phổ biến ở nhóm khách hàng có nguy cơ vỡ nợ cao, và kịp thời triển khai biện pháp phòng ngừa.

Bộ dữ liệu này bao gồm **255,347 dòng** và **18 cột thông tin**, chứa các đặc trưng về hồ sơ nhân khẩu học, thu nhập, lịch sử tín dụng và trạng thái hoàn trả của từng khoản vay.

Trong cuộc thi này, thí sinh sẽ sử dụng Power BI để trực quan hóa dữ liệu, phân tích xu hướng, nhận diện các yếu tố rủi ro và xây dựng **dashboard tương tác** hỗ trợ ra quyết định. Mục tiêu là giúp các bên liên quan dễ dàng hiểu được đặc điểm của các nhóm khách hàng, đánh giá rủi ro tín dụng và tối ưu hóa quy trình xét duyệt khoản vay.

Bảng chú giải dữ liệu

#	Tên biến	Kiểu dữ liệu	Mô tả
1	LoanID	Chuỗi	Mã định danh duy nhất của mỗi khoản vay
2	Age	Số nguyên	Tuổi của người vay
3	Income	Số nguyên	Thu nhập hàng năm của người vay
4	LoanAmount	Số nguyên	Số tiền vay
5	CreditScore	Số nguyên	Điểm tín dụng của người vay, phản ánh mức độ uy tín tài chính
6	MonthsEmployed	Số nguyên	Số tháng người vay đã làm việc tại nơi hiện tại
7	NumCreditLines	Số nguyên	Số lượng hạn mức tín dụng mà người vay đang có
8	InterestRate	Số thực	Lãi suất áp dụng cho khoản vay
9	LoanTerm	Số nguyên	Thời hạn vay (tháng)
10	DTIRatio	Số thực	Tỷ lệ nợ trên thu nhập của người vay (Debt-to-Income Ratio)

#	Tên biến	Kiểu dữ liệu	Mô tả
11	Education	Chuỗi	Trình độ học vấn cao nhất của người vay
12	EmploymentType	Chuỗi	Hình thức việc làm của người vay
13	MaritalStatus	Chuỗi	Tình trạng hôn nhân của người vay
14	HasMortgage	Chuỗi	Người vay có khoản vay thế chấp hay không
15	HasDependents	Chuỗi	Người vay có người phụ thuộc hay không
16	LoanPurpose	Chuỗi	Mục đích của khoản vay
17	HasCoSigner	Chuỗi	Khoản vay có người đồng bảo lãnh hay không
18	Default	Số nguyên	Biến mục tiêu: Khoản vay có bị vỡ nợ không (1 = Vỡ nợ, 0 = Không)

Lưu ý: Một số cột trong dataset có thể chứa giá trị bất thường, thí sinh được khuyến khích kiểm tra và xử lý dữ liệu trước khi xây dựng báo cáo phân tích.

B. Hướng dẫn kết nối dữ liệu

Cài đặt Power BI

Thí sinh cần cài đặt Power BI Desktop để làm bài thi.

→ Truy cập link sau để xem hướng dẫn chi tiết cài đặt:

https://datapot.vn/student-support-portal/#Doi_voi_Power_BI

Hướng dẫn kết nối dữ liệu từ SQL Server của Datapot

→ Truy cập link sau để xem hướng dẫn kết nối vào SQL Server:

https://datapot.vn/power-bi-moc-lab-1-prepare-data-in-power-bi-desktop/#Task_2_Lay_du_lieu_tu_SQL_Servernbsp

Thông tin kết nối server

Connection type: Microsoft SQL Server

Server: analytics-challenge-sql.datapot.edu.vn

Authentication: SQL Server Database

User name: nextgen_user

Password: Nextgen2025_01

Thông tin cuộc thi

Datapot phối hợp cùng các đối tác chuyên môn tổ chức cuộc thi **NextGen Analytics Challenge 2025**, với mục tiêu tạo ra một sân chơi thực tế để người học có cơ hội rèn luyện kỹ năng xử lý – phân tích – trực quan hóa dữ liệu thông qua các bài toán mô phỏng từ doanh nghiệp.

Cuộc thi kéo dài từ **tháng 7 đến tháng 12 năm 2025**, bao gồm **4 chặng thi chính**, tương ứng với **4 lĩnh vực phân tích**:

- Tài chính (Finance)
- Nhân sự (HR)
- Chuỗi cung ứng (Supply Chain) & Hậu cần (Logistics)
- Bán lẻ (Retail) & Người tiêu dùng (Consumer)

Thông tin chung về cuộc thi

- **Form nộp bài:** <https://forms.gle/c7JjRWhtNgbtYSyT6>
- **Website Datapot Challenge | Hall of Fame:** <https://datapot.vn/challenge-hall-of-fame/>

Kênh liên lạc – cập nhật thông tin chính thức

- **Fanpage Datapot:** <https://www.facebook.com/DatapotAnalytics>
- **Group Facebook DABI – Data Analytics and Business Intelligence Vietnam:**
<https://www.facebook.com/groups/nguoiламdata>

Kênh hỗ trợ chuyên môn

- **Zalo Group hỗ trợ Q&A:** <https://zalo.me/g/ekxslq015>
- **Kênh YouTube Datapot:** <http://info.datapot.vn/kenhYouTubeDatapot>
- **Tài nguyên hỗ trợ thí sinh:** <https://datapot.vn/tong-hop-tai-nguyen-ho-tro-thi-sinh-nextgen-analytics-challenge-2025/>