

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



Đồ án chuyên ngành - CO4029

Báo cáo

KHAI PHÁ DỮ LIỆU NGÂN HÀNG - TUẦN 5

Giảng viên hướng dẫn: PGS.TS Trần Minh Quang
ThS. Đỗ Thanh Thái

Sinh viên thực hiện: 2014486 - Đậu Xuân Thành
2014545 - Nguyễn Xuân Thắng

Mục lục

1. Tổng quan	3
2. Tìm kiếm & Thu thập dữ liệu (Data collection)	3
2.1. Tập dữ liệu 1 [1]	3
2.2. Tập dữ liệu 2 [2]	4
3. Xác định yêu cầu	4
3.1. Yêu cầu chức năng	4
3.2. Yêu cầu phi chức năng	5
4. Use-case diagram	7
4.1. Use-case diagram toàn hệ thống	7
4.2. Use-case diagram và đặc tả cho QUẢN LÝ ĐƠN VAY	8
5. Tổng kết và kế hoạch tuần tới	10
Tài liệu tham khảo	11

Danh mục hình ảnh

Hình 1: Sơ đồ use-case cho toàn hệ thống.	7
Hình 2: Sơ đồ use-case cho tính năng QUẢN LÝ ĐƠN VAY.	8

1. Tổng quan

Báo cáo tuần 5 của nhóm tập trung vào giải quyết những vấn đề theo sự hướng dẫn của GVHD (Thầy Quang), bao gồm:

- Tìm kiếm dữ liệu tin cậy về chủ đề ***Loan Approval Prediction*** và nghiên cứu, áp dụng những kỹ thuật khai phá dữ liệu lên đó.
- Bước đầu phân tích các yêu cầu chức năng, phi chức năng, vẽ sơ đồ use-case cho website của bộ công cụ.

2. Tìm kiếm & Thu thập dữ liệu (Data collection)

2.1. Tập dữ liệu 1 [1]

Mô tả: Tập dữ liệu được thu thập được các ngân hàng thu thập từ các khoản vay trước đây, nhằm mục đích xây dựng các mô hình dự đoán dựa trên các kỹ thuật khai phá dữ liệu, học máy. Để từ đó, phân loại các người đi vay xem họ có khả năng vỡ nợ hay không.

Thời gian thu thập dữ liệu: 2019

Độ lớn: hơn 140.000 dòng, với 34 thuộc tính.

Nguồn: Tập dữ liệu được giới thiệu bởi Kaggle - Một cộng đồng chia sẻ các nguồn dữ liệu đáng tin cậy, được nhiều bài báo về chủ đề dữ liệu, học máy sử dụng.

Các thuộc tính:

- *ID*: Mã số khoản vay.
- *year*: Năm dữ liệu được ghi nhận.
- *loan_limit*: Giới hạn khoản vay.
- *Gender*: Giới tính.
- *approv_in_adv*: Đã được phê duyệt trước.
- *loan_type*: Loại khoản vay.
- *loan_purpose*: Mục đích vay.
- *Credit_Worthiness*: Uy tín tín dụng.
- *open_credit*: Tín dụng mở. (Tín dụng xoay vòng)
- *business_or_commercial*: Kinh doanh hoặc thương mại
- *loan_amount*: Khoản vay.
- *rate_of_interest*: Lãi suất.
- *Interest_rate_spread*: Chênh lệch lãi suất (lãi suất cho vay - lãi suất tiền gửi).
- *Upfront_charges*: Lệ phí mà người đi vay phải trả trước khi khoản vay được chấp nhận.
- *term*: Kỳ hạn vay
- *Neg_ammortization*: Thỏa thuận trả góp hàng tháng
- *interest_only*: Khoản vay theo lịch trình (chỉ cần trả lãi)
- *lump_sum_payment*: Khoản vay được chia thành nhiều đợt trả hay không
- *property_value*: Giá trị tài sản.
- *construction_type*:
- *occupancy_type*:
- *Secured_by*:
- *total_units*:
- *income*: Thu nhập của người vay.
- *credit_type*: Loại tổ chức đánh giá điểm tín dụng
- *Credit_Score*: Điểm tín dụng, được tính dựa trên lịch sử trả nợ và hồ sơ tín dụng của người vay.
- *co-applicant_credit_type*: Thang điểm đánh giá tín dụng của người đi vay cùng (nếu có).

- *age*: Tuổi
- *submission_of_application*:
- *LTV*: loan-to-value là tỉ lệ khoản tiền thế chấp trên giá trị thẩm định của tài sản
- *Region*: Khu vực
- *Security_Type*:
- *Status*: Kết quả đánh giá.
- *dtir1*:

2.2. Tập dữ liệu 2 [2]

Mô tả: Tập dữ liệu từ LendingClub.com, tập dữ liệu được thu thập từ các khoản đầu tư trong quá khứ để phân tích, đánh giá, giúp cho các nhà đầu tư có thể có xác suất đầu tư vào các đối tượng có khả năng hoàn vốn cao mình.

Thời gian thu thập dữ liệu: 2007 - 2010

Độ lớn: hơn 9.578 dòng, với 14 thuộc tính.

Nguồn: Tập dữ liệu được thu thập từ LendingClub.com - Nền tảng giúp kết nối những người cần vay tiền với những người có tiền để cho vay (nhà đầu tư). Đây là nền tảng cung cấp các dịch vụ tài chính có trụ sở tại Mỹ, có uy tín cao trong ngành tài chính, đầu tư.

Các thuộc tính:

- *credit.policy*: 1 nếu người dùng đáp ứng các tiêu chí đánh giá của LendingClub.com, và ngược lại là 0.
- *purpose*: Mục đích của khoản vay (nhận các giá trị: *credit_card*, *debt_consolidation*, *educational*, *major_purchase*, *small_business*, and *all_other*)
- *int.rate*: Lãi suất khoản vay, người đi vay được LendingClub.com đánh giá có mức rủi ro cao sẽ có mức lãi suất cao hơn
- *installment*: Số tiền trả góp hàng tháng nếu khoản vay được chấp nhận.
- *log.annual.inc*: Nhật ký về thu nhập hàng năm tự sao kê của người đi vay.
- *dti*: debt-to-income: Tỷ lệ nợ trên thu nhập của người đi vay.
- *fico*: Điểm tín dụng FICO của người đi vay.
- *days.with.cr.line*: Số ngày người đi vay tín dụng.
- *revol.bal*: Số dư xoay vòng của người đi vay (số tiền chưa thanh toán ở mỗi kỳ thanh toán tín dụng).
- *revol.util*: Tỉ lệ sử dụng hạn mức tín dụng xoay vòng của người cho vay.
- *inq.last.6mths*: Số lượng yêu cầu vay của người đi vay trong 6 tháng vừa qua.
- *delinq.2yrs*: Số lần người đi vay quá hạn hơn 30 ngày trong vòng 2 năm qua.
- *pub.rec*: Số lượng hồ sơ công khai của người vay.

3. Xác định yêu cầu

3.1. Yêu cầu chức năng

YÊU CẦU CHUNG

ID	Feature	Description
G01	Login	User đăng nhập tài khoản vào hệ thống và được cấp tài nguyên theo vai trò. Ngoài ra còn có chức năng giúp người dùng đặt lại mật khẩu khi quên.

G02	Logout	Đăng xuất tài khoản khỏi hệ thống.
G03	Manage profile	Người dùng có thể xem/sửa các thông tin cơ bản của tài khoản như tên, email, ảnh đại diện, ...

HỆ THỐNG

ID	Feature	Description
S01	Notification	Hệ thống xử lý, hiển thị thông báo mới hoặc gửi email/sms cho người dùng mỗi khi có cập nhật về đơn vay hoặc các cập nhật khác.
S02	Model selection	Hệ thống phải có ít nhất 2 mô hình dự đoán cho người dùng lựa chọn.

DỰ ĐOÁN ĐƠN VAY

ID	Feature	Description
AL01	Create new loan	Admin có thể tạo mới đơn vay, bằng cách nhập các thông tin của ứng viên/hồ sơ theo form có sẵn, ứng với các biến trong mô hình dự đoán. Hoặc import dữ liệu từ file csv, json, ...
AL03	View loan detail	Admin có thể xem thông tin chi tiết của các đơn vay đã tạo.
AL04	Choose model	Admin có thể chọn mô hình dự đoán cho mỗi một đơn vay, các mô hình dự đoán có thể có như: Hồi quy logistic, Random forest, ...
AL05	View processed loan	Admin có thể xem lại các đơn hàng mình đã xử lý trong quá khứ. Các thông tin hiển thị bao gồm trạng thái (Accepted/Rejected), thời gian xử lý, kết quả dự đoán của mô hình.
AL06	Filter loan	Admin có thể tìm, lọc các đơn dựa trên các tiêu chí khác nhau như: đã xử lý, chưa xử lý, model dự đoán, ...
AL07	Manage loan	Admin có thể xóa/sửa các thông tin của các đơn vay chưa được xử lý như chỉnh sửa thông tin của các biến, chỉnh sửa mô hình dự đoán.

3.2. Yêu cầu phi chức năng

HIỆU SUẤT

ID	Feature	Description
----	---------	-------------

PE01		Các tác vụ xử lý dữ liệu (tính toán, dự đoán) yêu cầu trả kết quả trong vòng không quá 1 giây.
------	--	--

TÍNH TIN CẬY

ID	Feature	Description
RE01	Auditing	Hệ thống ghi lại logs khi có sự thay đổi tác động lên hệ thống hằng ngày.

THÂN THIỆN VỚI NGƯỜI DÙNG

ID	Feature	Description
ES01	Provide clear interface	Giao diện đẹp, rõ ràng, đảm bảo người dùng có thể tập trung vào những thông tin quan trọng, làm nổi bật những chỉ số, thông tin quan trọng bằng cách highlight, ...
ES02	Support multi platform	Trang web hỗ trợ giao diện responsive (tự động thay đổi các thông số về giao diện để phù hợp với kích thước màn hình) và đảm bảo tương thích trên nhiều hệ điều hành (iOS, Android, ...), trình duyệt khác nhau (Firefox, Chrome, MS Edge, ...)
ES03	Support multi language	Trang web hỗ trợ 2 ngôn ngữ: tiếng Anh và tiếng Việt
ES04	Provide clear feature	Trang web cung cấp các tính năng đơn giản, dễ thao tác, đảm bảo người dùng có thể thực hiện thành thạo sau 1 lần hướng dẫn

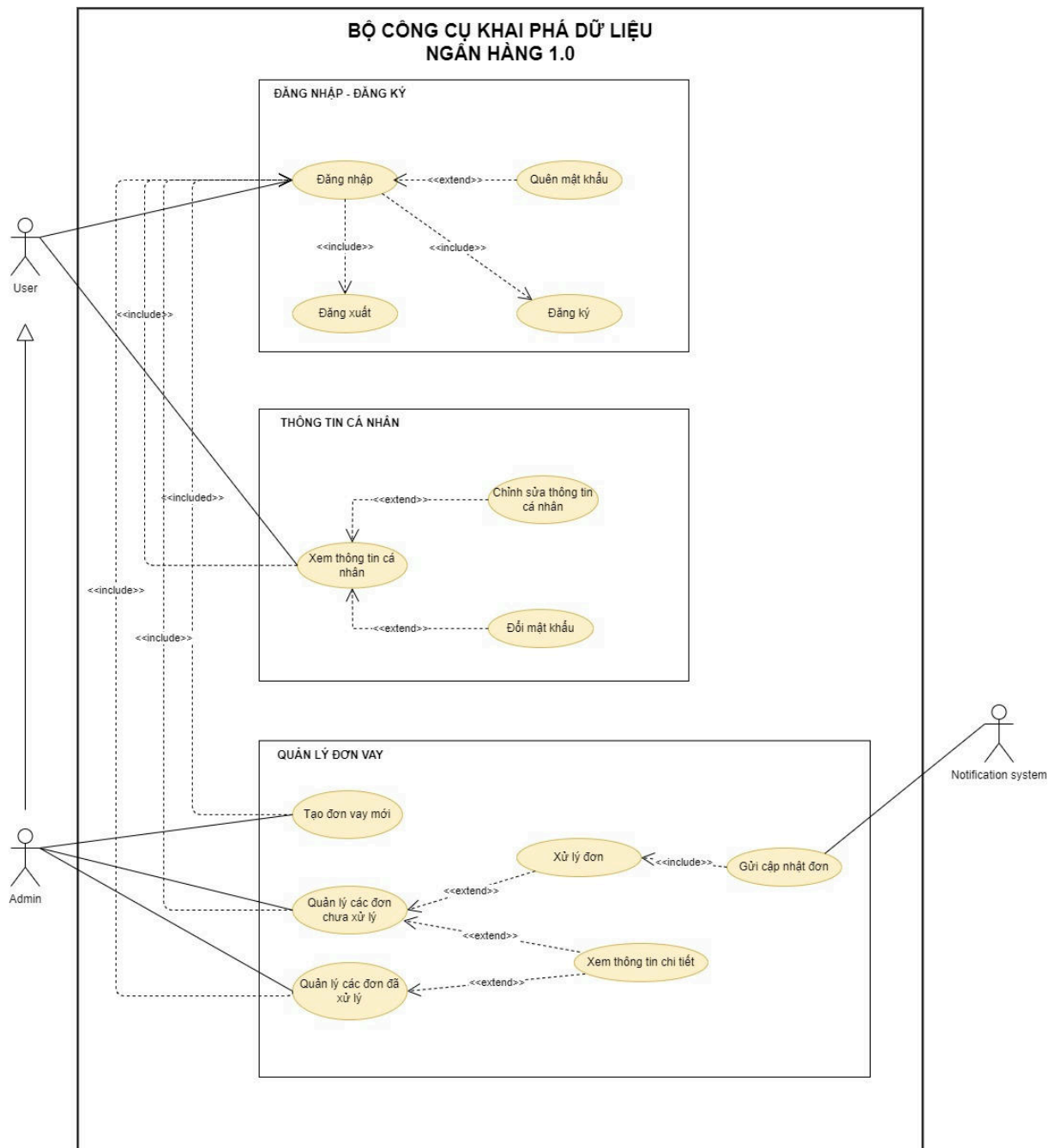
BẢO MẬT

ID	Feature	Description
SE01	Strong password	Mật khẩu của mỗi tài khoản được yêu cầu phải ít nhất 8 ký tự, bao gồm chữ cái in hoa và in thường, số và ký tự đặc biệt.
SE02	Auto block	Trong trường hợp người dùng đăng nhập thất bại 5 lần liên tiếp, hệ thống sẽ tự động khoá tài khoản, để mở khoá tài khoản, người dùng cần liên hệ với quản trị viên.
SE03	Secure website	Trang web sử dụng một số giao thức bảo mật trên mạng như TLS, SSL, JWT (Json Web Token) để bảo mật cho website.

4. Use-case diagram

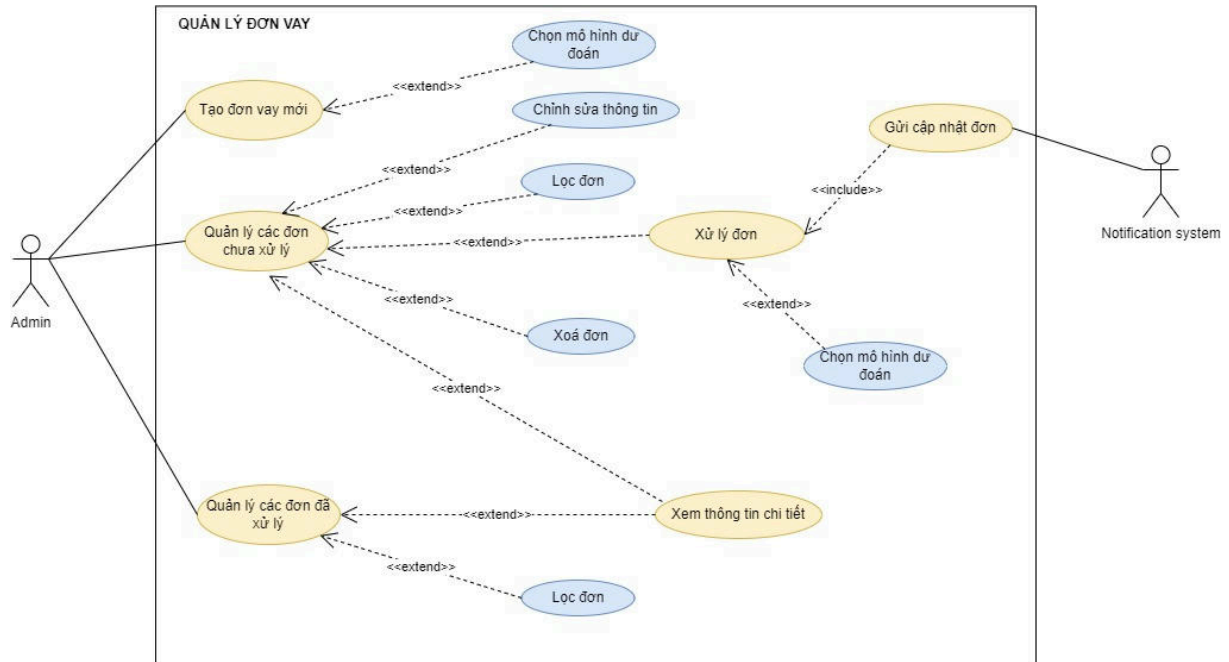
4.1. Use-case diagram toàn hệ thống

Xem chi tiết sơ đồ use-case của nhóm tại: <https://drive.google.com/file/d/1EtNV2Po7mtwW4zg2rioz8HqAY8JeO-W2/view?usp=sharing>



Hình 1: Sơ đồ use-case cho toàn hệ thống.

4.2. Use-case diagram và đặc tả cho QUẢN LÝ ĐƠN VAY



Hình 2: Sơ đồ use-case cho tính năng QUẢN LÝ ĐƠN VAY.

TẠO ĐƠN VAY MỚI

Use-case Name	Tạo đơn vay mới
Actors	Admin
Description	Admin tạo một đơn xin vay vốn mới dựa trên hồ sơ người dùng làm tại ngân hàng.
Trigger	Admin chọn vào button 'Tạo đơn mới' trong menu 'Quản lý đơn vay'.
Pre-Conditions	<ul style="list-style-type: none"> Admin đã đăng nhập thành công vào hệ thống và có quyền truy cập chức năng "Quản lý đơn vay".
Post-Conditions	Dữ liệu về đơn vay mới do admin thực hiện các thao tác được cập nhật vào trong cơ sở dữ liệu và hiển thị vào danh mục 'Đơn chưa được xử lý'
Normal Flow	<ol style="list-style-type: none"> Admin chọn mục 'Tạo đơn mới' trong menu 'Quản lý đơn vay'. Hệ thống mở cửa sổ 'Tạo đơn mới' bao gồm biểu mẫu yêu cầu nhập các thông tin cần thiết. Admin nhập các thông tin cần thiết vào biểu mẫu và nhấn 'Tạo'. Hệ thống xác nhận và thực hiện thêm thông tin của đơn mới vào cơ sở dữ liệu, đồng thời sử dụng model mà admin đã chọn để thực hiện dự đoán dựa trên các thông tin được cung cấp. Hệ thống thông báo tạo đơn mới thành công.
Alternative Flow	Không có

Exeption Flow	E1: Tại bước 2 2.1 Admin nhấn vào button ‘Huỷ’. 2.2 Hệ thống ẩn biểu mẫu tạo đơn mới. E2: Tại bước 4 4.1 Hệ thống xác định Admin chưa chọn loại mô hình để thực hiện dự đoán / Chưa nhập đủ các thông tin bắt buộc. 4.2 Hệ thống hiển thị warning và yêu cầu Admin thực hiện đủ các bước.
Constraints	Đơn mới tạo phải được nằm đầu tiên trong danh sách các đơn chưa xử lý.

QUẢN LÝ CÁC ĐƠN CHƯA XỬ LÝ

Use-case Name	Quản lý các đơn chưa xử lý
Actors	Admin
Description	Là một Admin, sẽ muốn xem xét các đơn đã tạo nhưng chưa được xử lý, từ đó có thể dễ dàng quản lý các đơn mới.
Trigger	Admin chọn vào tab ‘Đơn chưa xử lý’ trong menu ‘Quản lý đơn vay’.
Pre-Conditions	<ul style="list-style-type: none">Admin đã đăng nhập thành công vào hệ thống và có quyền truy cập chức năng “Quản lý đơn vay”.
Post-Conditions	Admin xem được thông tin của các đơn vay hiện có trong hệ thống tuy nhiên chưa được xử lý.
Normal Flow	<ol style="list-style-type: none">Admin chọn mục “Đơn chưa xử lý” trong menu “Quản lý đơn vay”.Hệ thống hiển thị danh sách các đơn vay có trong hệ thống nhưng có trạng thái “Chưa được xử lý”, được sắp xếp theo ngày tạo đơn. Mỗi đơn, ngoài các thông tin cơ bản, hệ thống hiển thị các thông tin quan trọng như trạng thái, ngày tạo, kết quả dự đoán của mô hình.
Alternative Flow	Không có
Exeption Flow	Không có
Constraints	Không có

XỬ LÝ ĐƠN

Use-case Name	Xử lý đơn
Actors	Admin
Description	Admin xử lý các đơn vay (Chấp nhận cho vay / Không chấp nhận)

Trigger	Admin chọn vào button “Xử lý” trong mỗi đơn vay chưa được xử lý trong tab “Đơn chưa xử lý” trong menu “Quản lý đơn vay”.
Pre-Conditions	<ul style="list-style-type: none">Admin đã đăng nhập thành công vào hệ thống và có quyền truy cập chức năng “Quản lý đơn vay”.
Post-Conditions	<ul style="list-style-type: none">Admin xử lý thành công đơn vay, trạng thái đơn vay thay đổi từ “Chưa được xử lý” thành “Được chấp nhận” hoặc “Bị từ chối”Dữ liệu được cập nhật vào cơ sở dữ liệu.Đơn vay chuyển từ tab “Đơn chưa xử lý” sang tab “Đơn đã xử lý”.
Normal Flow	<ol style="list-style-type: none">Admin nhấn vào button “Xử lý” của đơn vay bất kỳ trong danh sách được hiển thị trong tab “Đơn chưa xử lý”.Hệ thống hiển thị cửa sổ “Xử lý đơn + ID đơn” và hiển thị các thông tin cơ bản của đơn.Admin chọn vào button “Chấp nhận” hoặc “Từ chối” trong cửa sổHệ thống hiển thị thông báo xử lý đơn thành công.
Alternative Flow	A2: Tại bước 2 <ol style="list-style-type: none">Admin chọn button “Đổi mô hình dự đoán”.Hệ thống hiển thị các mô hình hiện có.Admin chọn mô hình mong muốn.Hệ thống tính toán, gửi kết quả dự đoán trở lại màn hình.
Exeption Flow	E1: Admin nhấn vào nút “Huỷ” Hệ thống ẩn cửa sổ đi, và trở lại giao diện của tab “Đơn chưa xử lý”.
Constraints	<ul style="list-style-type: none">Các đơn mới được xử lý sẽ được xuất hiện ở đầu của tab “Đơn đã xử lý”.

UPDATING ...

5. Tổng kết và kế hoạch tuần tới

Trong tuần, nhóm đã thực hiện phân tích các yêu cầu cho website và tìm kiếm các nguồn dữ liệu đáng tin cậy cho mô hình, cùng với đó phác thảo sơ đồ use-case cho hệ thống và đặc tả các use-case quan trọng.

Tuần tới, nhóm sẽ tiếp tục đặc tả các use-case còn lại để hoàn thiện phần phân tích hệ thống, đồng thời sẽ phác thảo cơ bản giao diện người dùng, cũng như chỉnh sửa theo sự hướng dẫn của GVHD.

Tài liệu tham khảo

- [1] M. Y. H, “Loan default dataset,” 2022. [Online]. Available: <https://www.kaggle.com/datasets/yasserh/loan-default-dataset/data>
- [2] ItsSuru, “Loan data,” 2021. [Online]. Available: <https://www.kaggle.com/datasets/itsuru/loan-data/data>