

CREDIT-AI-CONSENT

(Giải pháp tối ưu cho doanh nghiệp)

Các thành viên tham gia:

Nguyễn Thị Thanh Ngân
Nguyễn Thị Phương Thảo
Phạm Khánh Như
Lâm Chí Bảo
Mai Phú Hải

MỤC LỤC

| | |
|---|-----|
| MỤC LỤC..... | ii |
| DANH MỤC TỪ VIẾT TẮT..... | iii |
| CHƯƠNG 1. GIỚI THIỆU..... | 1 |
| CHƯƠNG 2. ĐIỂM MỚI & KHÁC BIỆT TRÊN THỊ TRƯỜNG..... | 2 |
| CHƯƠNG 3. GIẢI PHÁP CÔNG NGHỆ & MỨC ĐỘ HOÀN THIỆN (TECHNICAL EXECUTION & COMPLETION)..... | 3 |
| 3.1 Công nghệ sử dụng:..... | 13 |
| 3.2 Database..... | 14 |
| CHƯƠNG 4. Kết luận (Conclusion)..... | 15 |
| TÀI LIỆU THAM KHẢO..... | 17 |

DANH MỤC TỪ VIẾT TẮT

| STT | Chữ viết tắt/Kí hiệu | Định nghĩa Tiếng Việt/chú thích |
|-----|----------------------|--|
| 1 | <i>Consent</i> | Sự chấp thuận của người đi vay |
| 2 | <i>TxHash</i> | Chuỗi ký tự được tạo khi một giao dịch được ghi lại Blockchain |
| 3 | <i>Borrower</i> | Người đi vay |
| 4 | <i>CIC</i> | Trung tâm Thông tin tín dụng quốc gia Việt Nam |
| 5 | <i>metadata</i> | Dữ liệu về dữ liệu thô (raw data), chỉ là pattern của thông tin, không chứa các thông tin cá nhân nhạy cảm |

DANH MỤC HÌNH ẢNH

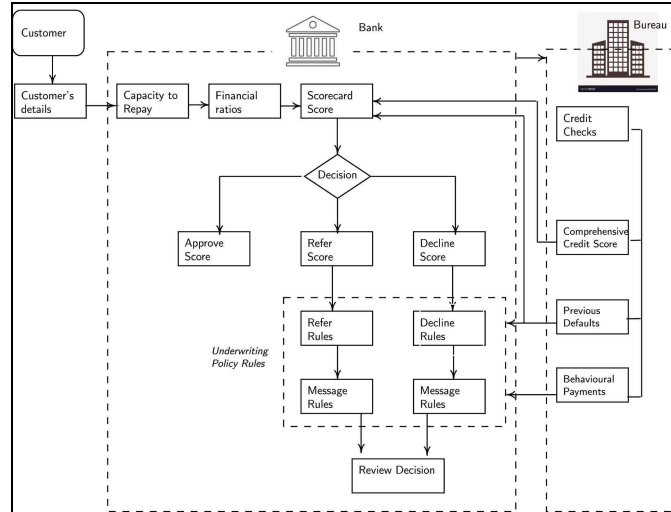
| | |
|---|----|
| Hình 1-1. Workflow hệ thống chấm điểm tín dụng thông thường | 2 |
| Hình 3-1. Blockchain system | 6 |
| Hình 3-2. Blockchain sequence | 7 |
| Hình 3-3. Blockchain Activity | 8 |
| Hình 3-4. Đồ thị thuật toán chữ ký | 8 |
| Hình 3-5. Thuật toán băm SHA256 | 9 |
| Hình 3-6. Thuật toán Boosting Gradient của LightGBM | 10 |
| Hình 3-7. Thuật toán Boosting Gradient của LightGBM | 10 |
| Hình 3-8. Scoring workflow (AI) | 12 |
| Hình 3-9. Approve metadata | 13 |
| Hình 3-10. Review metadata | 13 |
| Hình 3-11. Reject metadata | 14 |
| Hình 3-12. Hiệu năng mô hình LightGBM | 14 |
| Hình 3-13. Usecase của hệ thống | 16 |
| Hình 3-14. Database schema | 17 |

CHƯƠNG 1. GIỚI THIỆU

Dự án Intensifier hướng đến xây dựng công cụ chấm điểm tín dụng phi truyền thống, hỗ trợ cũng như mở rộng tầm nhìn và giải quyết các vấn đề mà các doanh nghiệp đang gặp phải. Trong hệ thống tài chính truyền thống, việc chấm điểm tín dụng của một khách hàng chủ yếu thông qua giao dịch, lịch sử vay nợ được lưu trong hệ thống Bureau vẫn được các ngân hàng tại Việt Nam tin dùng.

Tuy nhiên các tiêu chí trên vẫn còn tồn tại các mặt chủ quan, đặc biệt đối với các khách hàng tiềm năng chưa từng vay vốn chính thức do thiếu hồ sơ tín dụng hoặc tệp khách hàng muốn vay tiếp mà chưa đánh giá đúng hành vi, năng lực tài chính của họ tại thời điểm đó thông qua lịch sử trả nợ cũ. Chẳng hạn một sinh viên mới ra trường có công việc ổn định nhưng chưa từng sử dụng dịch vụ tín dụng sẽ khó tiếp cận khoản vay đầu tiên vì không có lịch sử tín dụng để tham chiếu. Hay nhiều tiểu thương, chủ hộ kinh doanh ở các vùng quê vốn chủ yếu giao dịch bằng tiền mặt, ít sử dụng thẻ ngân hàng, không có hồ sơ tín dụng rõ ràng, mặc dù duy trì hoạt động kinh doanh tốt và có khả năng trả nợ cao.

Điều này dẫn đến các hạn chế và rủi ro về quản trị trong bối cảnh nền kinh tế đang có nhiều biến động. Hiệp hội Ngân hàng Nhà nước Việt Nam (VNBA) phối hợp với Trung tâm Thông tin tín dụng quốc gia Việt Nam (CIC) đã tổ chức Tọa đàm kêu gọi tăng cường năng lực quản trị rủi ro và nâng cao chất lượng tín dụng nội bộ đã thấy được tầm quan trọng và cấp thiết của vấn đề trên.



Hình 1-1. Workflow hệ thống chấm điểm tín dụng thông thường

Trong bối cảnh công nghệ phát triển mạnh mẽ, đi kèm với các rủi ro lớn về an toàn thông tin cá nhân, tội phạm mạng xuất hiện khắp nơi đặc biệt trong ngành tài chính. Vụ việc CIC bị hacker tấn công cho thấy rủi ro lớn khi dữ liệu ứng dụng quốc gia tại một điểm, dễ trở thành mục tiêu khai thác và gây ảnh hưởng đến niềm tin người dùng. Song song với thách thức bảo mật, sự nổi trội về trí tuệ nhân tạo và phân tích dữ liệu đã tạo nhiều điều kiện mới nâng cao hiệu quả quy trình chấm điểm tín dụng, gia tăng độ tin cậy và góp phần cải thiện chất lượng danh mục tín dụng. Các tiêu chuẩn chấm điểm không dừng lại ở các thông tin tín dụng được cung cấp như trước mà còn chú trọng đến hành vi trực tuyến của khách hàng. Từ đó hình thành nên khái niệm chấm điểm tín dụng phi truyền thống, tức là khai thác các nguồn dữ liệu ngoài hệ thống ngân hàng như giao dịch thương mại điện tử, hoạt động mạng xã hội, thông tin viễn thông hay thói quen tìm kiếm và tiêu dùng trên Internet. Cách tiếp cận này mang lại cái nhìn đa chiều hơn về người vay, phản ánh không chỉ năng lực tài chính mà còn cả uy tín, thói quen và mức độ trách nhiệm trong hành vi của họ.

Phương pháp áp dụng cách chấm điểm phi truyền thống dựa vào khai thác dữ liệu hành vi người tiêu dùng trên mạng vốn đã quen thuộc trong bối cảnh tài chính toàn cầu. Hiện nay ở các ngân hàng và doanh nghiệp lớn ở VN bắt đầu thử nghiệm và chính thức áp dụng phương pháp này để đánh giá khách hàng. Tại nước ngoài phương pháp này đã áp dụng từ sớm và rộng rãi nhờ thị trường tài chính số phát triển và hệ thống pháp lý hoàn thiện. Tuy nhiên ở Việt nam còn

nhiều hạn chế, chủ yếu xuất hiện ở ngân hàng nhà nước Việt Nam và các doanh nghiệp lớn, chưa khai phá hết nguồn tiềm năng đến từ các tệp khách hàng khác nhau.

Nguyên nhân việc chấm điểm phi truyền thống chưa phổ biến ở Việt Nam là sự lo ngại về tính minh bạch và an toàn từ các ngân hàng đã làm rào cản khá lớn. Với tâm lý thận trọng của các doanh nghiệp, việc đề cao bảo mật là cấp thiết, và dự án của nhóm Intensifier sẽ chính là giải pháp cho các vấn đề mà các doanh nghiệp, thương nhân, cá nhân đang gặp phải.

CHƯƠNG 2. ĐIỂM MỚI & KHÁC BIỆT TRÊN THỊ TRƯỜNG

Đa nguồn phi truyền thống trong một khung thống nhất: SMS, danh bạ, mạng xã hội, e-commerce, web, điều này giúp chuẩn hóa thành schema chung để mở rộng dần. Intensifier quyết định áp dụng một nguồn dữ liệu phi truyền thống khác là **Email** để có góc nhìn tổng quan và tính điểm chính xác hơn chỉ các bộ dữ liệu phi truyền thống khác như nội dung tin nhắn SMS, danh bạ, hoạt động mạng xã hội, hành vi mua sắm e-commerce, xu hướng duyệt web.

Lý do:

Việc thu thập metadata của các bộ dữ liệu phi truyền thống như SMS, danh bạ, mạng xã hội, e-com, xu hướng duyệt web không phải là ngẫu nhiên, các bộ dữ liệu này sẽ thay nhau đóng vai trò làm chốt chặn chứng minh độ uy tín của bộ dữ liệu còn lại.

Việc áp dụng thêm Email làm metadata để chấm điểm tín dụng sẽ là giải pháp tương lai trong ngữ cảnh chuyển đổi số mạnh mẽ, đặc biệt là trong nước, với kỳ vọng đến năm 2030, 100% cơ quan, doanh nghiệp nhà nước và 70% doanh nghiệp tư nhân sẽ sử dụng dịch vụ điện toán đám mây (*Theo văn bản phê duyệt “Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030”*). Điều này cho thấy tương lai rộng mở của các dịch vụ trực tuyến và lưu trữ dữ liệu đã và đang trở thành xu hướng tất yếu đối với các doanh nghiệp. Vì vậy, việc giao tiếp thông qua internet gia tăng giữa các doanh nghiệp là điều rõ ràng.

Theo 1 báo cáo của Exploding Topics năm 2025 cho biết Google Workspace có thị phần hơn 50% trong thị trường phần mềm năng suất văn phòng, tức là một nửa của toàn bộ doanh nghiệp toàn cầu. Nên việc các dịch vụ online thông báo, giao tiếp thông qua Gmail chắc chắn sẽ gia tăng, nên nếu các doanh nghiệp cung cấp dịch vụ vay tín dụng “rào” được xu hướng tương lai này chắc chắn sẽ có được kết quả điểm tín dụng chính xác hơn rất nhiều trong việc phân tích xu hướng người đi vay. (*Theo Exploding Topics, “Google Workspace User Stats (2025)”*)

CHƯƠNG 3. GIẢI PHÁP CÔNG NGHỆ & MỨC ĐỘ HOÀN THIỆN (TECHNICAL EXECUTION & COMPLETION)

Credit-AI-consent là gì?

Credit-AI-consent là hệ thống tích hợp model AI để chấm điểm tín dụng, blockchain để minh bạch, xác thực và bảo đảm quyền lợi của khách hàng, hướng đến giải quyết các vấn đề của các doanh nghiệp cho vay cũng như là khách tệp khách hàng đang bị cho là “bỏ sót” trong thị trường.

MVP level (demo)

Blockchain (blockchain-lite):

Việc áp dụng công nghệ **blockchain** vào hệ thống chấm điểm tín dụng mang lại một giải pháp đột phá về bảo mật và minh bạch. Trong các hệ thống truyền thống, dữ liệu cá nhân nhạy cảm của người vay thường được lưu trữ tập trung, tạo ra nguy cơ cao bị rò rỉ hoặc tấn công từ các bên thứ ba.

Blockchain giải quyết vấn đề này bằng cách lưu trữ dữ liệu theo một cấu trúc phi tập trung và **bất biến**. Dữ liệu sẽ được mã hóa và lưu dưới dạng các giao dịch trong các **khối (blocks)**. Mỗi khối mới được thêm vào chuỗi sẽ chứa một **giá trị băm (hash)** của khối trước đó, tạo thành một chuỗi liên kết chặt chẽ.

Tính chất này đảm bảo rằng nếu một kẻ tấn công cố gắng thay đổi dữ liệu trong một khối đã có, giá trị băm của khối đó sẽ thay đổi. Điều này sẽ làm mất hiệu lực của tất cả các khối theo sau, và hệ thống sẽ ngay lập tức phát hiện và từ chối thay đổi đó. Nhờ vậy, blockchain cung cấp một số cái kỹ thuật số an toàn, minh bạch và không thể bị giả mạo, giúp bảo vệ dữ liệu người dùng một cách hiệu quả trong quá trình chấm điểm tín dụng.

Ở đây Intensifier xây dựng một hệ thống “blockchain-lite”, đã giản lược đi các tính năng heavy weight để demo dự án phù hợp với mức MVP. Hệ thống blockchain-lite này vẫn sẽ đáp ứng được tính minh bạch và an toàn cơ bản cho một hệ thống chấm điểm tín dụng để lưu trữ và chia sẻ các thông tin nhạy cảm của người đi vay.

Tính năng:

Quyền kiểm soát của người vay: Người vay (Borrower) chủ động cấp hoặc thu hồi quyền sử dụng các dữ liệu phi truyền thống, phục vụ mục đích chấm điểm tín dụng.

Minh bạch và truy vết: Mỗi quyết định được gắn với một Mã đồng thuận (ConsentId) và Mã giao dịch (TxHash) để phục vụ mục đích kiểm chứng lại về sau.

An toàn: Bất kỳ thay đổi nào đều để lại dấu vết, quyền đã thu hồi hoặc hết hạn sẽ ngay lập tức chặn chấm điểm.

Luồng hoạt động:

GIVE → STATUS → PROOF → REVOKE → STATUS → VERIFY

GIVE: cấp quyền truy cập dữ liệu, cho phép hệ thống sử dụng theo phạm vi và thời hạn do người đi vay set.

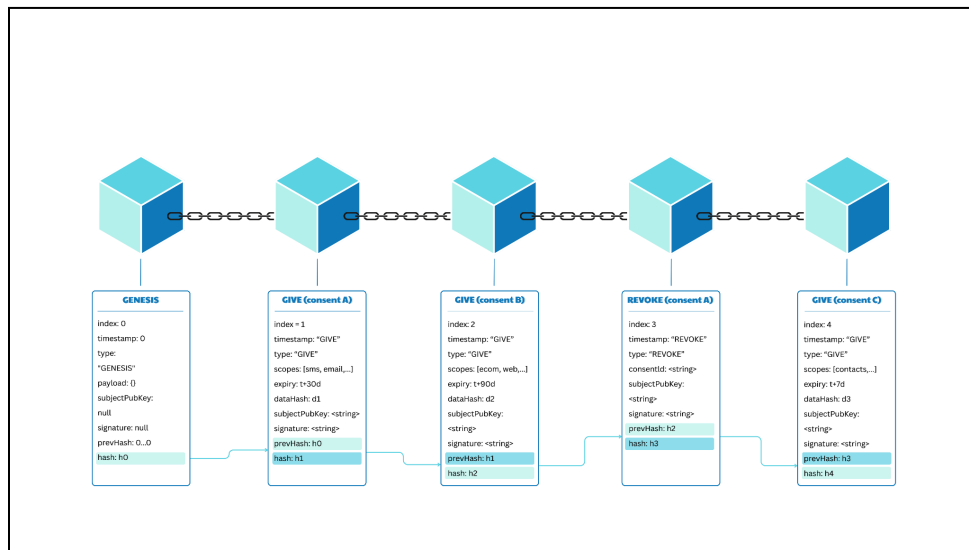
Xác thực consent khi chấm điểm: đảm bảo chỉ chấm điểm khi người đi vay còn cho phép.

STATUS: cung cấp trạng thái về hiệu lực của quyền do người đi vay set.

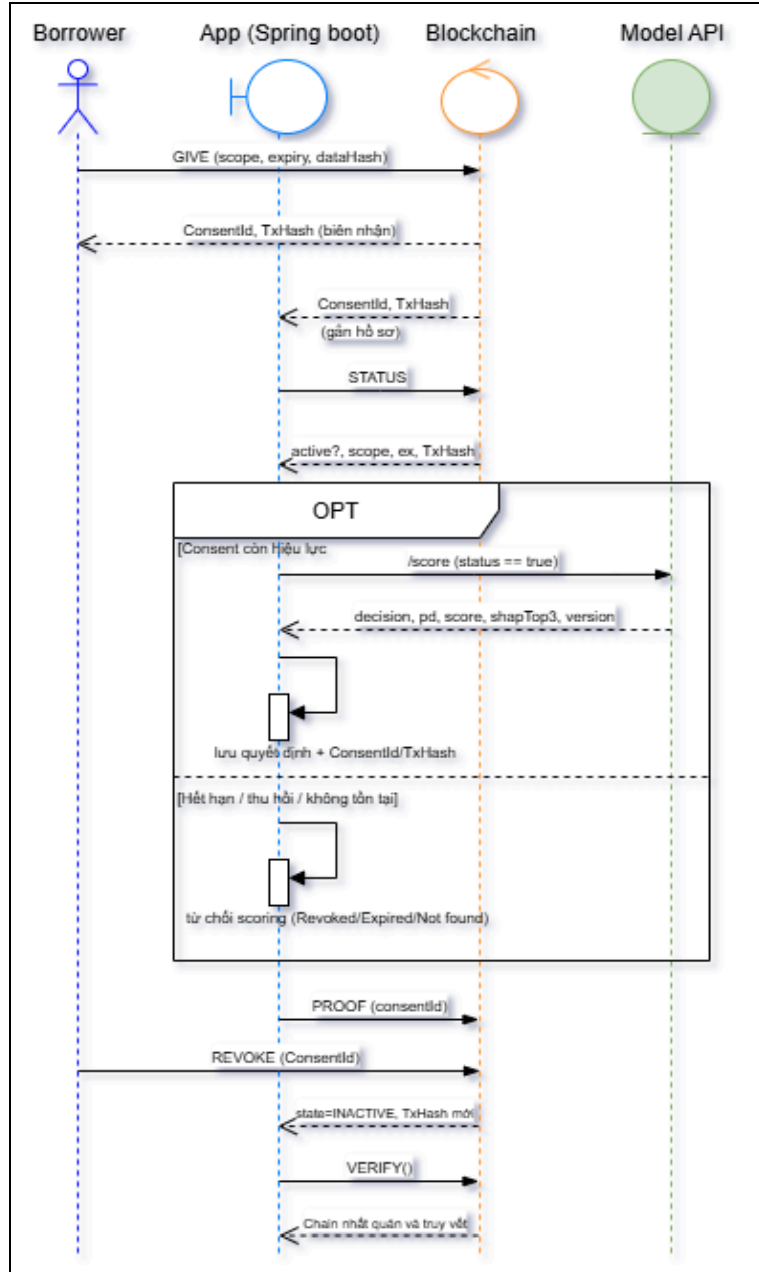
REVOKE: cho phép người đi vay có quyền dừng ngay việc sử dụng dữ liệu.

PROOF: cung cấp bằng chứng các hành động đã diễn ra liên quan tới một consent.

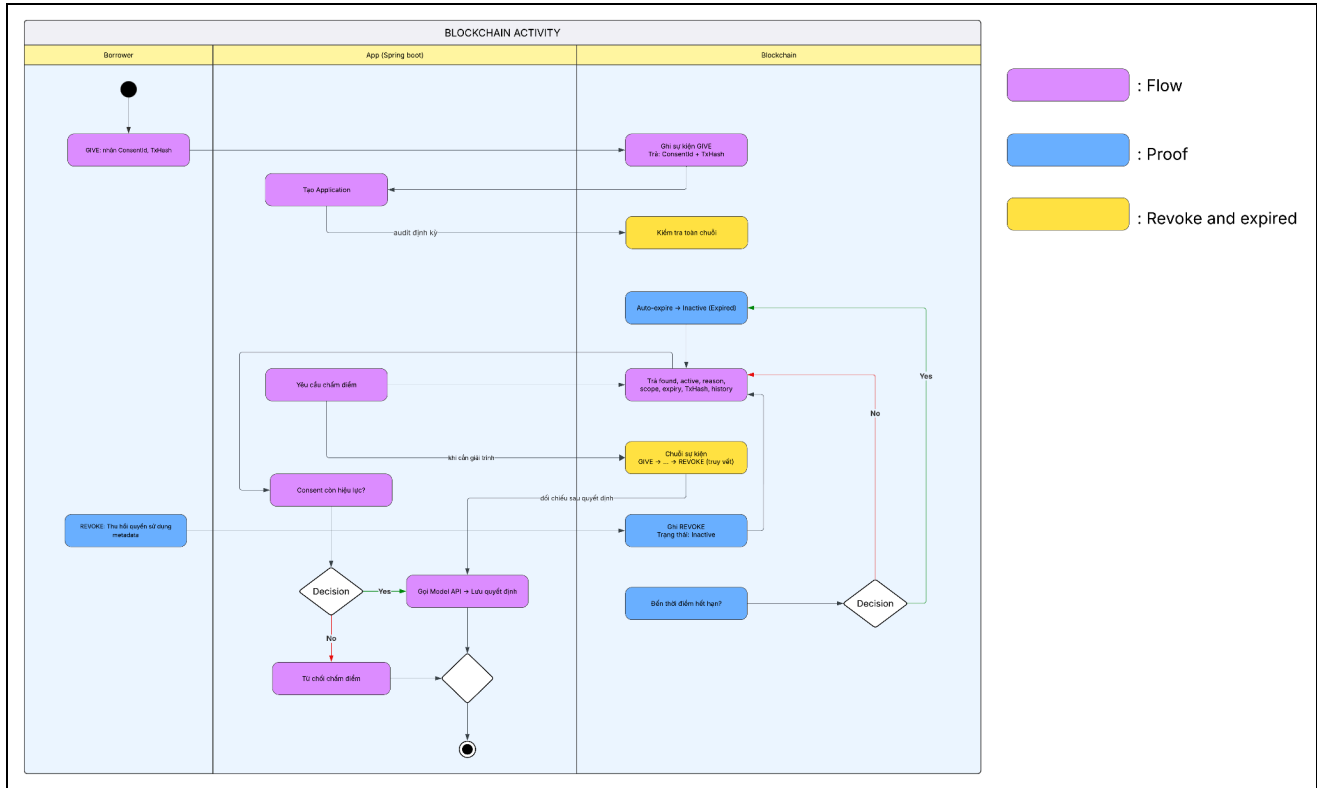
HEALTH/VERIFY: tự kiểm tra sức khỏe hệ thống để phát hiện bất thường.



Hình 3-1. Blockchain system




Hình 3-2. Blockchain sequence

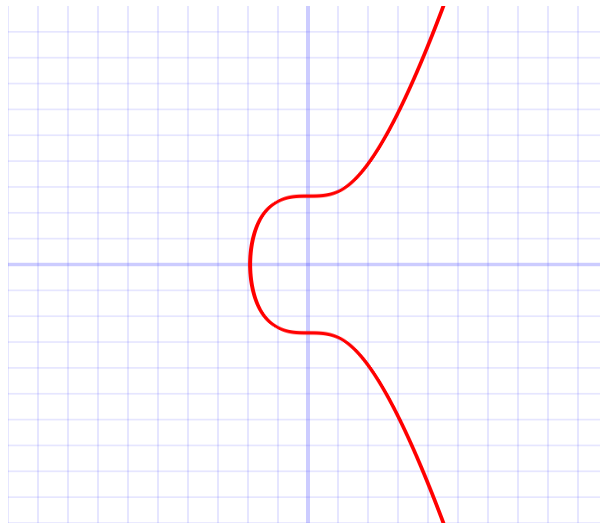


Hình 3-3. Blockchain Activity

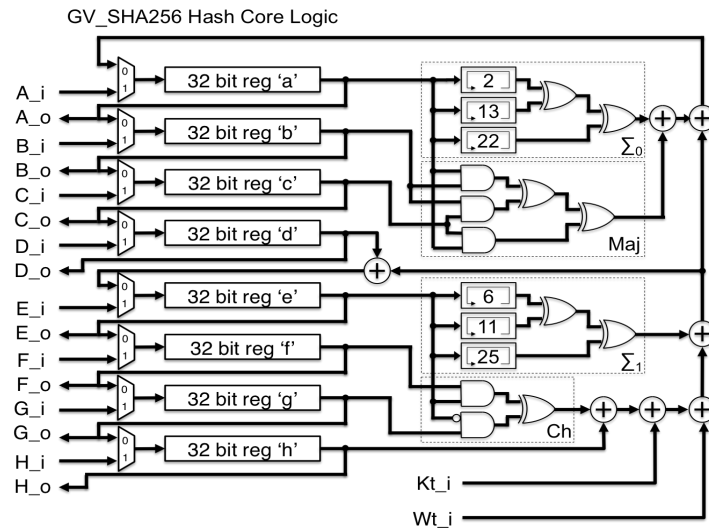
Công nghệ sử dụng:

 Node.js (Express)

Thuật toán chữ ký nhóm sử dụng **secp256k1**: $y^2 = x^3 + 7 \pmod{p}$



Hình 3-4. Đồ thị thuật toán chữ ký

Thuật toán băm: **SHA256**

Hình 3-5. Thuật toán băm SHA256

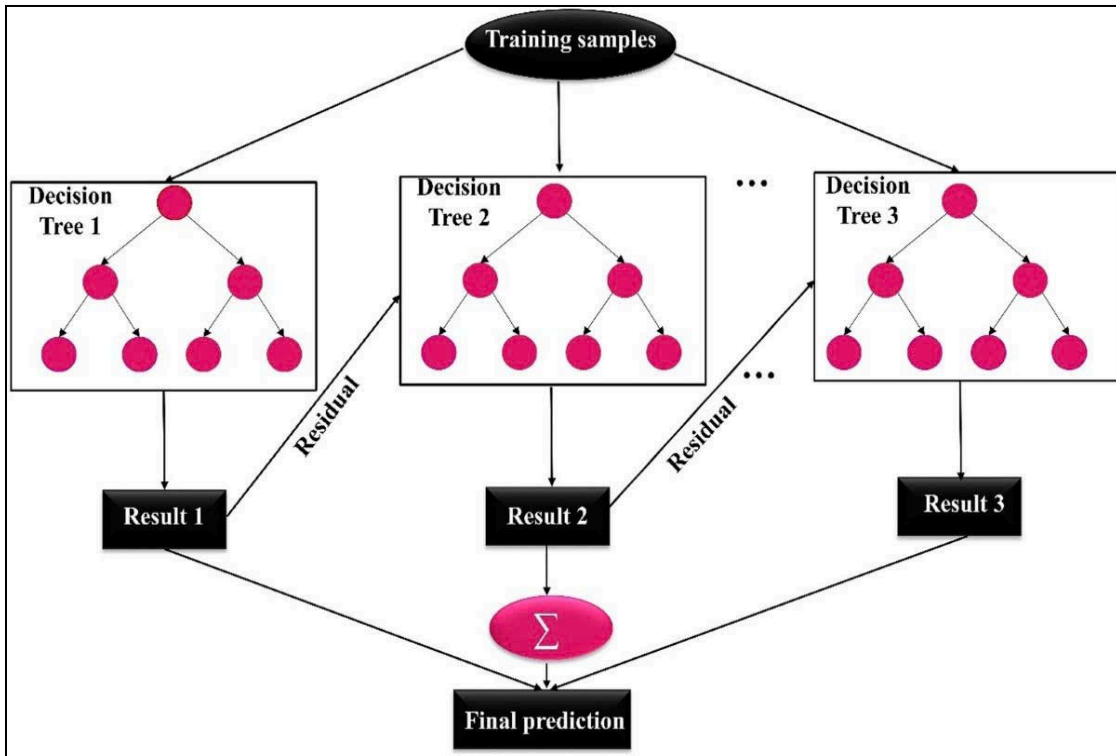
AI (model LightGBM)

Dự án Intensifier sử dụng model **LightGBM**, là một thuật toán học máy Gradient Boosting, nổi bật với khả năng xử lý khối lượng dữ liệu lớn và tốc độ huấn luyện nhanh. Với cơ chế **leaf-wise + histogram**, LightGBM tập trung học sâu vào những đặc trưng quan trọng giúp tiết kiệm bộ nhớ vừa nâng cao độ chính xác.

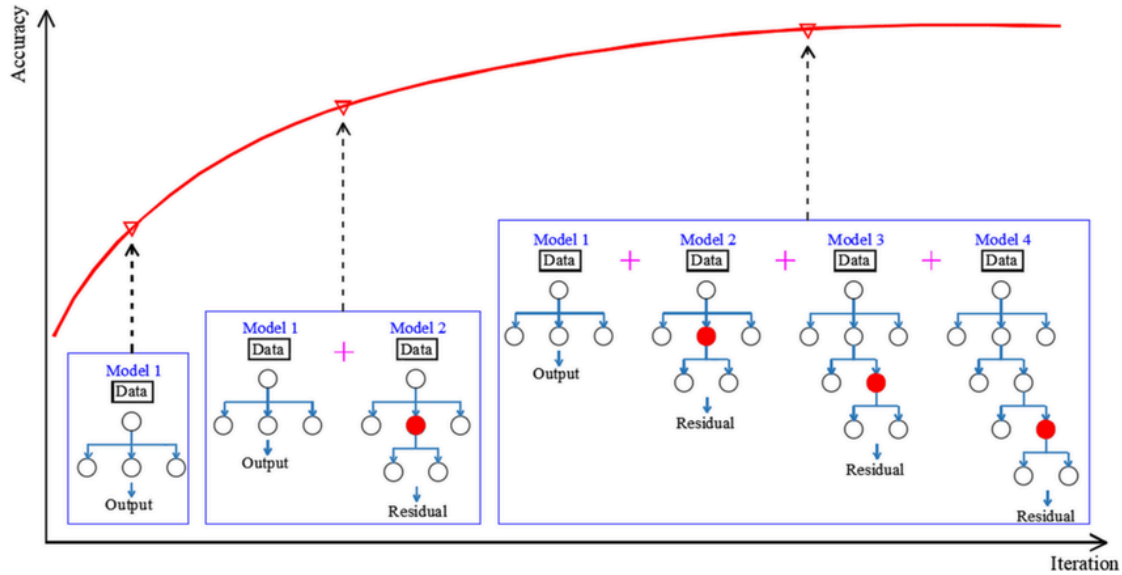
| Tiêu chí | AdaBoost | XGBoost | LightGBM |
|----------------------------|------------|--------------------|--|
| Tốc độ huấn luyện | Chậm | Nhanh hơn AdaBoost | Nhanh nhất (nhờ leaf-wise + histogram) |
| Độ chính xác | Trung bình | Cao | Rất cao (tối ưu lỗi tốt hơn) |
| Khả năng xử lý dữ liệu lớn | Hạn chế | Tốt | Rất tốt (thiết kế cho scale) |
| Tiêu thụ bộ nhớ | Nhiều | Vừa phải | Ít nhờ binning (histogram) |
| Hỗ trợ song song/GPU | Không | Có | Có, tối ưu mạnh |

| | | | |
|--------------------------------|------------------------|---------------------|---------------------------------------|
| Ứng dụng thực tế | Bài toán nhỏ | Kaggle, data mining | Chăm điểm tín dụng, dự đoán, xếp hạng |
| Độ chính xác thực tế (tabular) | Trung bình, nhạy nhiều | Cao | Rất cao với tham số phù hợp |

Thuật toán Gradient Boosting chính là yếu tố chính dẫn đến quyết định sử dụng model này.



Hình 3-6. Thuật toán Boosting Gradient của LightGBM



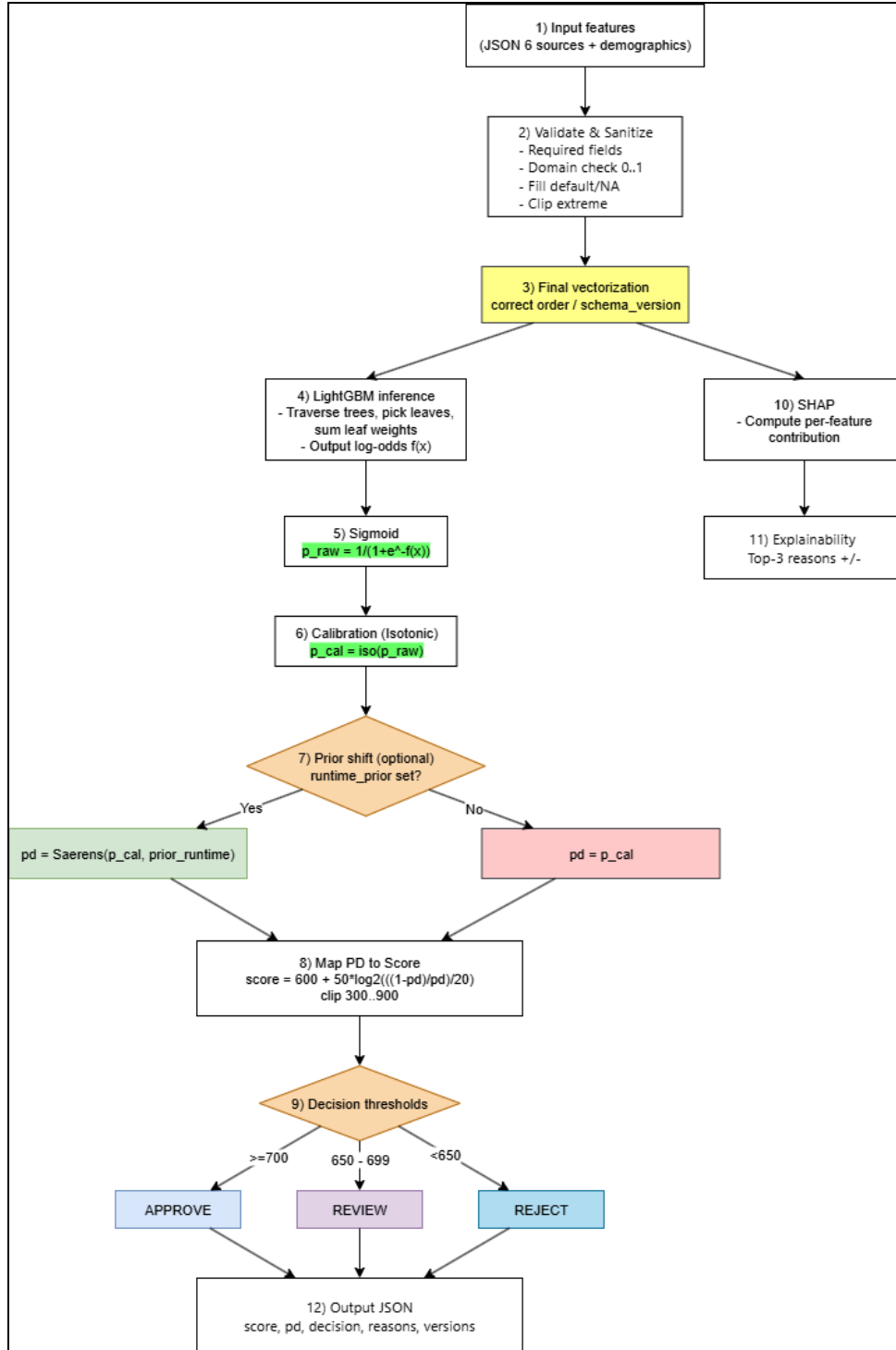
Hình 3-7. Thuật toán Boosting Gradient của LightGBM

Xử lý dữ liệu đa dạng: Gradient Boosting có khả năng khai thác cả dữ liệu truyền thống (lịch sử vay, CIC) và phi truyền thống như hành vi mua sắm online, dữ liệu viễn thông, sử dụng ví điện tử, hay thậm chí dữ liệu mạng xã hội.

Tăng độ chính xác: Bằng cách học từ sai số của các cây trước đó, mô hình dần cải thiện chất lượng dự đoán, giảm thiểu tình trạng bỏ sót khách hàng tiềm năng.

Công bằng hơn: Không còn phụ thuộc hoàn toàn vào lịch sử tín dụng, mô hình có thể đánh giá năng lực chi trả dựa trên thói quen tài chính và hành vi thực tế.

Ra quyết định nhanh: Nhờ khả năng huấn luyện và suy luận nhanh, Gradient Boosting giúp ngân hàng đưa ra quyết định cấp tín dụng kịp thời, hỗ trợ mở rộng dịch vụ tài chính đến nhiều nhóm khách hàng hơn.



Hình 3-8. Scoring workflow (AI)

Ta sẽ demo 3 bộ metadata với 3 kết quả kèm 3 các tín hiệu dẫn đến mức % pd, dưới đây là giải thích cho các tín hiệu làm biến động điểm.

PD: Xác suất khách hàng không trả được nợ (vỡ nợ).
 PD thấp → khách hàng an toàn, ít rủi ro. PD ↓
 PD cao → khách hàng nguy hiểm, dễ vỡ nợ. PD ↑



*Thông tin cơ bản

"pd_true": 0,
 "age": 38,
 "monthly_income_vnd": 60000000, "emp_contract": 1, "emp_formal": 1,
 "emp_gig": 0, "emp_self": 0, "emp_student": 0,
 "region_CT": 0, "region_DN": 0, "region_HCM": 1, "region_HN": 0, "region_HP": 0,
 "region_NT": 0,
 "gender_F": 0, "gender_M": 1,
 *Dữ liệu liên lạc và mạng xã hội

*Thương mại điện tử

"ecom_orders": 36, "ecom_spend_sum": 3000.0, "ecom_basket_avg": 83.3,
 "ecom_cod_ratio": 0.00, "ecom_return_ratio": 0.00,
 "ecom_cat_beauty_ratio": 0.05, "ecom_cat_electronics_ratio": 0.25,
 "ecom_cat_fashion_ratio": 0.10, "ecom_cat_gaming_ratio": 0.00,
 "ecom_cat_grocery_ratio": 0.20, "ecom_cat_home_ratio": 0.25,
 "ecom_cat_others_ratio": 0.15,

*Hoạt động web và email

"web_visits": 80, "web_banking_ratio": 0.40, "web_education_ratio": 0.15,
 "web_entertainment_ratio": 0.10, "web_gambling_ratio": 0.00,
 "web_news_ratio": 0.20, "web_shopping_ratio": 0.10, "web_short_loan_ratio":
 0.00, "web_social_ratio": 0.05, "email_count": 180, "email_overdue_ratio":
 0.00, "email_type_job_ratio": 0.10, "email_type_promo_ratio": 0.05,
 "email_type_statement_ratio": 0.15, "email_type_transaction_ratio": 0.55,
 "email_type_unknown_ratio": 0.05, "email_type_utility_ratio": 0.10

Nam, 38 tuổi - độ tuổi ổn định
 Hồ Chí Minh - Khu vực thành thị lớn, điều kiện kinh tế xã hội tốt
 60 triệu/ tháng - thu nhập cao PD ↓
 Công việc hợp đồng chính thức - nghề nghiệp ổn định

Có kết nối xã hội thật, giao tiếp tốt, không spam

Mạng lưới rộng, danh bạ không có liên hệ có nguy cơ rủi ro, quan hệ đa dạng, ổn định PD ↓

Mạng xã hội mạnh, hoạt động thường xuyên, không tài khoản ảo, không có hành vi vi phạm PD ↓

Mua sắm đều, tiêu dùng khá lớn, mức chi tiêu trung bình hợp lý

Danh mục đa dạng, chi tiết thiết yếu, tài sản giá trị cho thấy mức sống cao PD ↓

Thường xuyên dùng ngân hàng online, quen thuộc với tài chính
 Không dính web cờ bạc, tin dụng đen

Không nợ quá hạn PD ↓

Chủ yếu giao dịch thật, tài chính minh mạch

Kết luận: được duyệt nhờ thu nhập cao, công việc ổn định hành vi tài chính minh bạch, mạng xã hội lành mạnh, không có dấu hiệu nợ quá hạn

Hình 3-9. Approve metadata



*Thông tin cơ bản

"pd_true": 0,
 "age": 34,
 "monthly_income_vnd": 45000000, "emp_contract": 0, "emp_formal": 1,
 "emp_gig": 0, "emp_self": 0, "emp_student": 0,
 "region_CT": 0, "region_DN": 0, "region_HCM": 1, "region_HN": 0, "region_HP": 0,
 "region_NT": 0,
 "gender_F": 0, "gender_M": 1,
 *Dữ liệu liên lạc và mạng xã hội
 "sms_count": 200, "sms_in_ratio": 0.6,
 "sms_fin_ratio": 0.0, "sms_kw_ratio": 0.0,
 "contacts_count": 350, "contacts_risky_ratio": 0.0,
 "rel_coworker_ratio": 0.25, "rel_family_ratio": 0.25,
 "rel_friend_ratio": 0.20, "rel_service_ratio": 0.10, "rel_unknown_ratio": 0.20,
 "social_rows": 60, "social_active_days": 25, "social_posts_sum": 20,
 "social_likes_sum": 300, "social_friends_avg": 500, "social_violations_avg": 0,
 "social_engagement": 0.15,
 *Thương mại điện tử
 "ecom_orders": 24, "ecom_spend_sum": 1200.0, "ecom_basket_avg": 50.0,
 "ecom_cod_ratio": 0.05, "ecom_return_ratio": 0.02,
 "ecom_cat_beauty_ratio": 0.05, "ecom_cat_electronics_ratio": 0.20,
 "ecom_cat_fashion_ratio": 0.10, "ecom_cat_gaming_ratio": 0.05,
 "ecom_cat_grocery_ratio": 0.20, "ecom_cat_home_ratio": 0.20,
 "ecom_cat_others_ratio": 0.20,
 *Hoạt động web và email
 "web_visits": 60, "web_banking_ratio": 0.30, "web_education_ratio": 0.20,
 "web_entertainment_ratio": 0.10, "web_gambling_ratio": 0.0,
 "web_news_ratio": 0.20, "web_shopping_ratio": 0.20, "web_short_loan_ratio":
 0.0, "web_social_ratio": 0.20, "email_count": 120, "email_overdue_ratio": 0.0,
 "email_type_job_ratio": 0.10, "email_type_promo_ratio": 0.10,
 "email_type_statement_ratio": 0.10, "email_type_transaction_ratio": 0.60,
 "email_type_unknown_ratio": 0.0, "email_type_utility_ratio": 0.10

Nam, 34 tuổi - độ tuổi ổn định
 Hồ Chí Minh - Khu vực thành thị lớn, điều kiện kinh tế xã hội tốt
 45 triệu/ tháng - thu nhập cao **PO+**
 Công việc chính thức **PO+**
 Nhưng không có hợp đồng ghi nhận **PO-**

Có kết nối xã hội thật, giao tiếp tốt, không spam **PO+**

Mạng lưới khá rộng, danh bạ không có liên hệ có nguy cơ rủi ro, quan hệ đa dạng, ổn định **PO+**

Ghi nhận có số lạ gọi tới **PO+**

Mạng xã hội khá, hoạt động thường xuyên, không tài khoản ảo, không có hành vi vi phạm **PO+**

Mua sắm đều, tiêu dùng tương đối, mức chi tiêu trung bình hợp lý **PO+**

Danh mục đa dạng, cân bằng chi tiêu thiết yếu và tiêu khiển

Có sử dụng ngân hàng online

Không dính web cờ bạc, tin dụng đen **PO+**

Không nợ quá hạn **PO+**

Chủ yếu giao dịch thật, tài chính minh mạch **PO+**

Kết luận: Thu nhập cao có công việc ổn định, có quản lý chi tiêu. Mạng xã hội lành mạnh, không có dấu hiệu nợ quá hạn. Tuy nhiên công việc không có hợp đồng ghi nhận. Có khả năng được Approve

Hình 3-10. Review metadata



*Thông tin cơ bản

"pd_true": 0,
 "age": 22,
 "monthly_income_vnd": 3000000, "emp_contract": 0, "emp_formal": 0,
 "emp_gig": 1, "emp_self": 0, "emp_student": 0,
 "region_CT": 1, "region_DN": 0, "region_HCM": 0, "region_HN": 0, "region_HP": 0,
 "region_NT": 0,
 "gender_F": 0, "gender_M": 1,
 *Dữ liệu liên lạc và mạng xã hội
 "sms_count": 80, "sms_in_ratio": 0.2,
 "sms_fin_ratio": 0.4, "sms_kw_ratio": 0.4,
 "contacts_count": 60, "contacts_risky_ratio": 0.35,
 "rel_coworker_ratio": 0.05, "rel_family_ratio": 0.05,
 "rel_friend_ratio": 0.10, "rel_service_ratio": 0.20, "rel_unknown_ratio": 0.60,
 "social_rows": 60, "social_active_days": 5, "social_posts_sum": 60,
 "social_likes_sum": 10, "social_friends_avg": 80, "social_violations_avg": 2,
 "social_engagement": 0.02,
 *Thương mại điện tử
 "ecom_orders": 8, "ecom_spend_sum": 120.0, "ecom_basket_avg": 15.0,
 "ecom_cod_ratio": 0.90, "ecom_return_ratio": 0.30,
 "ecom_cat_beauty_ratio": 0.10, "ecom_cat_electronics_ratio": 0.0,
 "ecom_cat_fashion_ratio": 0.50, "ecom_cat_gaming_ratio": 0.20,
 "ecom_cat_grocery_ratio": 0.0, "ecom_cat_home_ratio": 0.0,
 "ecom_cat_others_ratio": 0.20,
 *Hoạt động web và email
 "web_visits": 90, "web_banking_ratio": 0.0, "web_education_ratio": 0.0,
 "web_entertainment_ratio": 0.40, "web_gambling_ratio": 0.20,
 "web_news_ratio": 0.0, "web_shopping_ratio": 0.40, "web_short_loan_ratio":
 0.70, "web_social_ratio": 0.50, "email_count": 40, "email_overdue_ratio": 0.20,
 "email_type_job_ratio": 0.0, "email_type_promo_ratio": 0.70,
 "email_type_statement_ratio": 0.0, "email_type_transaction_ratio": 0.10,
 "email_type_unknown_ratio": 0.10, "email_type_utility_ratio": 0.10

Nam, 22 tuổi - độ tuổi trẻ, thường chưa ổn định sự nghiệp
 Cần Thơ
 3 triệu/ tháng - thu nhập thấp, khả năng trả nợ hạn chế **PO+**
 Công việc thời vụ - nghề nghiệp tự do, không ổn định **PO+**

Lượng tin nhắn ít, chủ yếu gửi đi

Nhiều tin liên quan đến tài chính/ vay mượn, số liên hệ ít, có nguy cơ rủi ro **PO+**

Số lạ gọi tới nhiều **PO+**

Mạng lưới xã hội yếu, thiếu ổn định **PO+**

Có hành vi vi phạm **PO+**

Giao dịch nhỏ, không thường xuyên **PO+**

Chỉ tiêu về tiêu khiển, không phải hàng thiết yếu **PO+**

Không sử dụng ngân hàng online

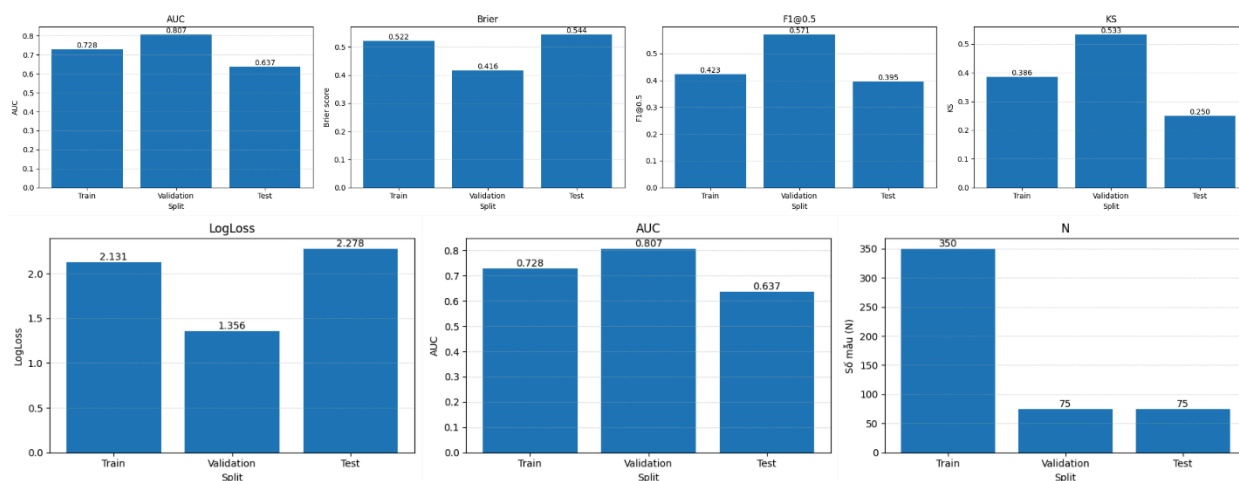
Truy cập nhiều vào trang web cờ bạc, vay nóng **PO+**

Có nợ quá hạn **PO+**

Phần lớn là email quảng cáo, ít giao dịch thật **PO+**

Kết luận: Chứa nhiều yếu tố rủi ro cao. Thu nhập và công việc không ổn định. Liên lạc và mạng xã hội yếu. Hành vi thương mại điện tử không minh bạch. Tiếp xúc nhiều với vay nóng và cờ bạc, kèm nợ quá hạn

Hình 3-11. Reject metadata



Hình 3-12. Hiệu năng mô hình LightGBM

Web (framework Spring MVC-java)

Trong dự án Intensifier, nhóm sử dụng mô hình Spring MVC, một mô hình web end-to-end cho số đông các trang web đang hiện hành trên internet ngày nay.

Vì sao sử dụng Spring MVC?

Để hoàn thiện dự án với mức độ MVP (Minimum Viable Product), thì mô hình Spring MVC với framework Spring Boot là lựa chọn hoàn hảo để xây dựng cho hệ thống này.

Phân lớp rõ ràng (Model - View - Controller), giúp dòng nghiệp vụ dễ được kiểm soát hơn: Controller → Service → Repository → Database.

Đây là mô hình được tích hợp hệ sinh thái chuẩn bao gồm JSON và thời gian được cấu hình sẵn qua Jackson → ta có được payload nhất quán giữa input và output.

Tích hợp sẵn RestTemplate để gọi Model API (dịch vụ blockchain-lite).


Mô hình Spring MVC có cấu trúc nhiều endpoint nhỏ → dự án dễ build CRUD và gọi hỗ trợ từ JdbcTemplate → dễ mở rộng dự án đến mức độ hoàn thiện mà không cần phải refactor quá nhiều code.


Dễ dàng tích hợp AI và Blockchain:


AI (Model API): Hệ thống sẽ xác định consent của người đi vay có hợp lệ hay không, sau đó gọi FastAPI để lấy các tham số PD (Probability of default), điểm, minh bạch 3 lý do cho người dùng rủi ro (SHAP), sau đó sẽ ghi kết quả vào Database để hiển thị và truy vết.

Blockchain-lite: Backend khi chấm điểm sẽ đối chiếu txHash mới nhất của consent; kết quả chấm điểm luôn gắn với consentId/txHash để truy vết. Mọi quyết định đều chứng minh được là chỉ xảy ra khi còn consent hợp lệ.

3.1 Công nghệ sử dụng:


 **Spring Boot + Spring Web MVC:** khởi động nhanh, routing REST quen thuộc, dễ cấu hình, cộng đồng lớn → phù hợp MVP cần tốc độ triển khai. (Cấu trúc *controller/service/repository/config*).


 **Jackson (JSON + JavaTime):** đảm bảo định dạng thời gian/JSON thống nhất, tránh lỗi timezone.

 **JdbcTemplate/SimpleJdbcCall:** gọi SP trực tiếp, kiểm soát SQL và hiệu năng; giảm ma sát khi schema đã có.

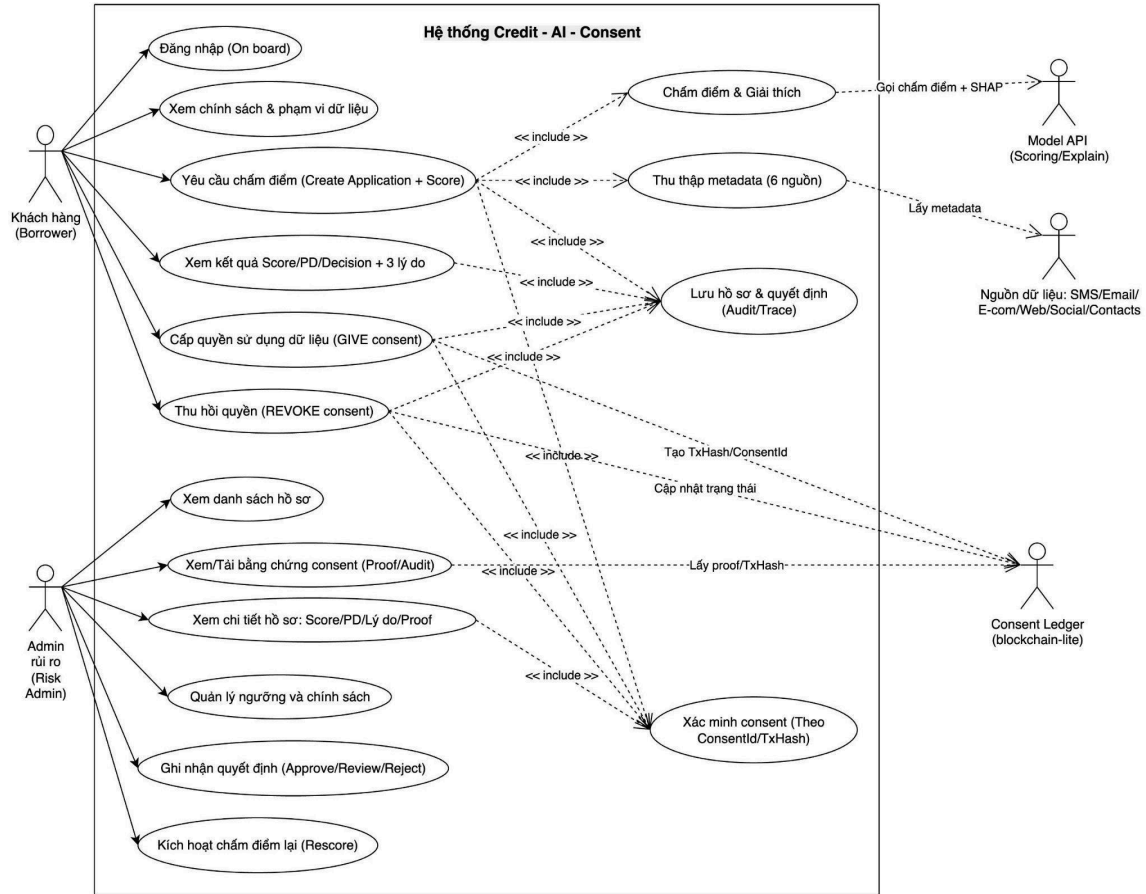
Lombok: giảm boilerplate (constructor/getter/setter) → code service/repo gọn.

RestTemplate client: tích hợp AI (FastAPI) và Ledger qua HTTP; đơn giản, đủ dùng cho dự án mức độ MVP.

 **OpenAPI/Swagger:** để tự sinh tài liệu API khi cần demo.

 **Thymeleaf + TailwindCSS (CDN) + Vanilla JS (fetch) cho mini-UI:** dùng làm giao diện Borrower và Risk, gọi trực tiếp API và trình bày dưới dạng JSON.

Use case của hệ thống:



Hình 3-13. Usecase của hệ thống

3.2 Database

Intensifier sử dụng hệ cơ sở dữ liệu SQL Server để lưu trữ cũng như phát triển trong tương lai.

Dưới đây là Schema Database của hệ thống Credit-AI-Consent.

CHƯƠNG 4. KẾT LUẬN (CONCLUSION)

Dự án hướng đến hoàn thiện 100% các tính năng của dự án, chứ không chỉ ở mức MVP bao gồm:

SDK cho mobile và web: Nhóm sẽ phát triển các bộ công cụ phần mềm (SDK) cho Android, iOS và Web để xử lý các thao tác dữ liệu cơ bản, kèm theo một hệ thống quản lý việc thu thập dữ liệu qua hàng đợi tin nhắn (như Kafka hoặc RabbitMQ).

Feature store và quản lý dữ liệu: Nhóm sẽ triển khai một hệ thống lưu trữ dữ liệu có cấu trúc cho việc quản lý tính năng, với các tính năng như kết hợp dữ liệu theo thời gian và khả năng cập nhật dữ liệu theo từng bước hoặc bổ sung dữ liệu đã thiếu.

Cải thiện công bằng: Nhóm sẽ triển khai các công cụ giám sát công bằng nâng cao, chẳng hạn như một bảng điều khiển để kiểm tra sự bất công trong quyết định, mẫu thông báo cho người dùng khi có quyết định bất công và báo cáo chi tiết về hiệu suất của mô hình.

Kiểm thử mô hình và tự động quay lại: Dự án sẽ hỗ trợ A/B testing với nhiều phiên bản mô hình và khả năng tự động quay lại phiên bản trước khi có lỗi, giúp việc thử nghiệm mô hình an toàn hơn trong môi trường sản xuất.

MLOps tự động hóa: Quy trình sẽ tự động tái huấn luyện mô hình theo lịch trình, kết hợp kiểm tra chất lượng dữ liệu và gửi thông báo qua Slack hoặc email, cùng với một hệ thống quản lý mô hình tập trung để theo dõi tất cả các phiên bản.

Bảo mật nâng cao: Bảo mật sẽ được tăng cường với việc quản lý khóa mã hóa (KMS), tự động xoay vòng khóa, ghi nhật ký ngăn ngừa rò rỉ dữ liệu (DLP) và một cơ sở dữ liệu riêng biệt cho thông tin cá nhân nhạy cảm (PII).

Quản lý đồng ý nâng cao: Nhóm sẽ cung cấp khả năng kiểm soát chi tiết hơn đối với dữ liệu đồng ý, bao gồm việc thiết lập thời gian hết hạn cho các nguồn dữ liệu cụ thể và khả năng thu hồi đồng ý trên toàn bộ hệ thống với đầy đủ báo cáo kiểm toán.

Mở rộng nguồn dữ liệu: Nhóm sẽ tích hợp thêm các nguồn dữ liệu như hóa đơn tiện ích, ví điện tử và khám phá việc sử dụng học liên kết (FL) hoặc bảo mật riêng tư phân tán (DP) để tăng cường bảo vệ dữ liệu và quyền riêng tư.

Giao diện người dùng (UI/UX): Một giao diện người dùng hoàn chỉnh sẽ được xây dựng, bao gồm ứng dụng cho người vay và bảng điều khiển quản lý rủi ro cho quản trị viên, với các tính năng chi tiết cho phân tích nhóm và so sánh.

TÀI LIỆU THAM KHẢO

- [1] Báo Nhân Dân. (n.d.). *Ứng dụng mô hình xếp hạng tin dụng hiện đại*. Truy cập từ <https://nhandan.vn/ung-dung-mo-hinh-xep-hang-tin-dung-hien-dai-post897521.html>
- [2] Cổng Thông tin điện tử Chính phủ. (2025, 11 tháng 9). *Hệ thống văn bản*. Truy cập từ <https://vanban.chinhphu.vn/?pageid=27160&docid=200163>
- [3] Báo Tiền Phong. (2025, 11 tháng 9). *Hacker tấn công Trung tâm Thông tin Tín dụng Quốc gia*. Truy cập từ <https://tienphong.vn/hacker-tan-cong-trung-tam-thong-tin-tin-dung-quoc-gia-post1777489.tpo>
- [4] Báo Điện tử Chính phủ. (2025, 11 tháng 9). *CIC bị hacker tấn công, VNCERT khuyến cáo người dân nâng cao cảnh giác*. Truy cập từ <https://baochinhphu.vn/cic-bi-hacker-tan-cong-vncert-khuyen-cao-nguoi-dan-nang-cao-ca-nh-giac-102250911201710342.htm>
- [5] Exploding Topics. (2025, 5 tháng 6). *Google Workspace User Stats (2025)*. Truy cập từ <https://explodingtopics.com/blog/google-workspace-stats>
- [6] Myriantous, G. (2024, 11 tháng 12). *LightGBM: A Guide*. Built In. Truy cập từ <https://builtin.com/articles/lightgbm>
- [7] GeeksforGeeks. (2025, 15 tháng 7). *LightGBM (Light Gradient Boosting Machine)*. Truy cập từ <https://www.geeksforgeeks.org/machine-learning/lightgbm-light-gradient-boosting-machine/>
- [8] Top10hcmc-city. (n.d.). *Gradient Boost là gì? Tầm quan trọng và ứng dụng*. Truy cập từ <https://top10hcmc-city.com/gradient-boost-la-gi-tam-quan-trong-va-ung-dung/>
- [9] InterData. (n.d.). *LightGBM là gì? A-Z về Light Gradient Boosting Machine nên biết*. Truy cập từ <https://interdata.vn/blog/lightgbm-la-gi/>

- [10] ichi.pro. (n.d.). *Hiểu về lightgbm*. Truy cập từ <https://ichi.pro/vi/hieu-ve-lightgbm-131033555708922>