HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Dương Bảo Anh

XÂY DỰNG HỆ KHUYẾN NGHỊ HỖ TRỢ KHÁCH HÀNG SỬ DỤNG CÁC DỊCH VỤ NGÂN HÀNG

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI – 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Dương Bảo Anh

XÂY DỰNG HỆ KHUYẾN NGHỊ HỖ TRỢ KHÁCH HÀNG SỬ DỤNG CÁC DỊCH VỤ NGÂN HÀNG

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04

ĐỀ ÁN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. NGUYỄN DUY PHƯƠNG

HÀ NỘI - 2025

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong đề án tốt nghiệp là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Hà Nội, ngày tháng 02 năm 2026

Tác giả đề án

Dương Bảo Anh

LỜI CẢM ƠN

Lời đầu tiên, em xin bày tỏ lòng biết ơn sâu sắc và chân thành nhất tới Quý Thầy, Cô giáo – những người đã trực tiếp tham gia giảng dạy và truyền đạt kiến thức trong suốt quá trình em học tập tại Học viện Công nghệ Bưu chính Viễn thông, đặc biệt là các Thầy, Cô thuộc Khoa Công nghệ Thông tin 1 và Khoa Đào tạo sau đại học. Những kiến thức nền tảng và chuyên sâu mà Thầy, Cô trang bị là yếu tố tiên quyết giúp em có thể thực hiện công trình nghiên cứu này.

Đặc biệt, em xin gửi lời tri ân sâu sắc nhất tới Tiến sĩ Nguyễn Duy Phương, người đã nhận lời hướng dẫn cho em. Thầy không chỉ định hướng cho em một đề tài nghiên cứu có ý nghĩa khoa học và thực tiễn, mà còn dành nhiều thời gian, tâm huyết để tận tình hướng dẫn, chỉ bảo, đồng hành và tạo mọi điều kiện thuận lời trong suốt quá trình em thực hiện đề án. Sự hỗ trợ chuyên môn sâu sắc và những lời khuyên quý báu của Thầy là động lực và kim chỉ nam giúp em hoàn thành đề án này.

Em cũng xin gửi lời cảm ơn tới gia đình, bạn bè và đồng nghiệp đã luôn động viên, khích lệ và chia sẻ khó khăn, tạo điều kiện tốt nhất để em tập trung hoàn thành chương trình học tập và nghiên cứu của mình.

Em đã nỗ lực hết sức, làm việc nghiêm túc và vận dụng những kiến thức, phương pháp luận đã được học để hoàn thành đề án. Tuy nhiên, do còn hạn chế về kinh nghiệm nghiên cứu cũng như giới hạn về mặt thời gian, đề án chắc chắn không thể tránh khỏi những thiếu sót và hạn chế.

Kính mong Quý Thầy, Cô đóng góp những ý kiến quý báo để đề án của em được hoàn thiện hơn nữa.

Em xin trân trọng cảm ơn!

MỤC LỤC

LỜI CAM ĐOAN i
LỜI CẢM ƠN ii
MỤC LỤC iii
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT v
DANH MỤC CÁC BẢNG vi
DANH MỤC HÌNH ẢNH vii
MỞ ĐẦU 1
1. Lý do chọn đề tài 1
2. Tổng quan về vấn đề nghiên cứu 2
3. Mục đích nghiên cứu 3
4. Đối tượng và phạm vi nghiên cứu 3
5. Phương pháp nghiên cứu 4
NỘI DUNG 5
CHƯƠNG 1: GIỚI THIỆU VỀ HỆ KHUYẾN NGHỊ 5
1.1. Tổng quan về hệ khuyến nghị 5
1.1.1. Khái niệm hệ khuyến nghị 5
1.1.2. Phân loại hệ khuyến nghị 6
1.1.3. Cấu trúc tổng thể của hệ khuyến nghị 14
1.2. Tổng quan về dịch vụ ngân hàng 17
1.2.1. Phân loại dịch vụ ngân hàng cho khách hàng cá nhân 17
1.2.2. Phân loại nhu cầu và hành vi người dùng 19
1.2.3. Vai trò của hệ khuyến nghị trong lĩnh vực tài chính – ngân hàng. 20
1.3. Mô hình khuyến nghị trong lĩnh vực tài chính – ngân hàng 20
1.4. Kết luận chương 1 22
CHƯƠNG 2: HỆ KHUYẾN NGHỊ CÁC DỊCH VỤ NGÂN HÀNG 24
2.1. Phân tích người dùng 24
2.2. Nguồn dữ liệu và đặc điểm 26
2.3. Phân tích yêu cầu hệ thống khuyến nghị 28
2.4. Thiết kế kiến trúc hệ thống khuyến nghị 30

2.5. Phương pháp ước lượng và dự đoán 32
2.5.1. Phương pháp dựa trên luật (Rule – based System).....32
2.5.2. Phương pháp lọc theo nội dung (Content – based Filtering).....33
2.5.3. Phương pháp lọc cộng tác (Collaborative Filtering) 33
2.5.4. Mô hình lai (Hybrid Recommendation Model).....33
2.6. Kết luận chương 2 34
CHƯƠNG 3: XÂY DỰNG ỨNG DỤNG.....36
3.1. Nền tảng và công cụ triển khai 36
3.1.1. Các công nghệ và thư viện sử dụng.....36
3.1.2. Cơ sở lựa chọn công nghệ 37
3.2. Xây dựng tập dữ liệu mẫu.....37
3.2.1. Hồ sơ khách hàng (Customers).....38
3.2.2. Danh mục sản phẩm ngân hàng (Products) 40
3.2.3. Lịch sử giao dịch (Transactions) 41
3.3. Xử lý và huấn luyện dữ liệu.....44
3.3.1. Tiền xử lý dữ liệu 44
3.3.2. Huấn luyện mô hình khuyến nghị.....45
3.3.3. Đánh giá mô hình 51
3.4. Phát triển giao diện ứng dụng 52
3.4.1. Gợi ý cho khách hàng hiện hữu.....52
3.4.2. Gợi ý cho khách hàng mới.....52
3.4.3. Luồng xử lý tổng thể.....53
3.5. Thử nghiệm và đánh giá hệ thống.....54
3.5.1. Kết quả thử nghiệm 54
3.5.2. Đánh giá và phân tích 55
3.6. Kết luận chương 3 56
KẾT LUẬN.....57
DANH MỤC CÁC TÀI LIỆU THAM KHẢO.....60

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Ký hiệu, chữ viết tắt	Tên tiếng anh đầy đủ	Tên tiếng việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
CBF	Content – based Filtering	Lọc dựa trên nội dung
CF	Collaborative Filtering	Lọc cộng tác
KPDL		Khai phá dữ liệu
MAE	Mean Absolute Error	Sai số tuyệt đối trung bình
NDCG	Normalized Discounted Cumulative Gain	Mức tăng tích lũy giảm dần chuẩn hóa
RB	Rule - based	Dựa trên luật
RMSE	Root Mean Square Error	Căn bậc hai của sai số bình phương trung bình.
TF – IDF	Term Frequency – Inverse Document Frequency	Tần suất – Nghịch đảo tần suất tài liệu
TTNT		Trí tuệ nhân tạo

DANH MỤC CÁC BẢNG

Bảng 3.1: Các công nghệ và thư viện sử dụng.....36
Bảng 3.2: Kết quả đánh giá mô hình.....52

DANH MỤC HÌNH ẢNH

Hình 1.1: Cấu trúc tổng thể của hệ khuyến nghị.....15
Hình 2.1: Luồng hoạt động tổng quát của hệ khuyến nghị.....32
Hình 3.1: Sơ đồ quan hệ giữa các bảng trong tập dữ liệu mô phỏng.....38
Hình 3.2: Pipeline huấn luyện mô hình Hybrid.....46
Hình 3.3: Sơ đồ trình tự của hệ thống gợi ý.....54
Hình 3.4: Demo kết quả gợi ý.....55

MỞ ĐẦU

1. Lý do chọn đề tài

Trong giai đoạn ngành tài chính – ngân hàng đẩy mạnh chuyển đổi số, nhu cầu cá nhân hóa dịch vụ và tối ưu hóa trải nghiệm khách hàng ngày càng trở nên quan trọng và không thể thiếu. Ngày nay, các ngân hàng không chỉ dừng lại ở những sản phẩm truyền thống như tiền gửi hay vay tiêu dùng, mà còn mở rộng sang các dịch vụ đa kênh, ngân hàng số và nhiều giải pháp tài chính phức tạp. Chính sự đa dạng này khiến khách hàng – nhất là những khách hàng thiếu hiểu biết về tài chính – gặp nhiều khó khăn trong việc lựa chọn sản phẩm phù hợp với nhu cầu của bản thân.

Bên cạnh đó, các hệ thống khuyến nghị ngày càng khẳng định vai trò quan trọng trong việc phân tích hành vi, hiểu nhu cầu và cung cấp các đề xuất mang tính cá nhân hóa cho người dùng. Một số nghiên cứu chỉ ra rằng các phương pháp lọc theo nội dung (Content – based Filtering) và lọc cộng tác (Collaborative Filtering) đạt hiệu quả cao trong việc dự đoán sở thích của khách hàng [3], [4], [5], [7]. Ngoài ra, các hệ thống lai (Hybrid Recommender Systems) cho thấy khả năng khắc phục hạn chế của từng phương pháp đơn lẻ, từ đó cải thiện độ chính xác và mức độ bao phủ. Các mô hình này đặc biệt hiệu quả trong nhiều lĩnh vực, trong đó có tài chính – ngân hàng [6], [8].

Song song với đó, hệ chuyên gia dựa trên luật (Rule – based Expert System) vẫn giữ vai trò quan trọng trong những bài toán đòi hỏi khả năng giải thích rõ ràng và minh bạch trong quá trình đưa ra quyết định, đặc biệt trong những hệ thống mà nghiệp vụ được quy định bằng các quy tắc chặt chẽ như lĩnh vực ngân hàng. Việc kết hợp giữa mô hình khuyến nghị hiện đại và cơ chế suy diễn dựa trên luật giúp tạo ra giải pháp vừa đáp ứng nhu cầu theo từng khách hàng, vừa đảm bảo tính minh bạch – yếu tố thiết yếu trong dịch vụ tài chính.

Dựa trên các yêu cầu thực tế và định hướng nghiên cứu đã được phân tích, đề tài "Xây dựng hệ khuyến nghị hỗ trợ khách hàng sử dụng các dịch vụ ngân hàng" được lựa chọn với mục tiêu phát triển một hệ thống gợi ý thông minh, giúp người dùng nhanh chóng xác định được sản phẩm phù hợp, đồng thời hỗ trợ ngân hàng tối

ưu hóa việc cung ứng dịch vụ và đẩy mạnh quá trình chuyển đổi số trong công tác chăm sóc khách hàng.

Để trình bày nội dung một cách hệ thống, khoa học và nhất quán, đề án được tổ chức thành ba chương chính như sau:

Chương 1: Giới thiệu về hệ khuyến nghị

Chương này trình bày cơ sở lý thuyết về hệ khuyến nghị, bao gồm khái niệm, phân loại, cấu trúc hệ thống, cùng tổng quan về dịch vụ ngân hàng và hành vi người dùng, làm nền tảng cho các chương tiếp theo.

Chương 2: Hệ khuyến nghị các dịch vụ ngân hàng

Chương này tập trung phân tích dữ liệu người dùng, xác định yêu cầu, thiết kế kiến trúc và lựa chọn phương pháp dự đoán để xây dựng mô hình hệ khuyến nghị cho dịch vụ ngân hàng.

Chương 3: Xây dựng ứng dụng

Chương này mô tả quá trình triển khai hệ khuyến nghị thành ứng dụng thực nghiệm, bao gồm xử lý dữ liệu, tích hợp mô hình dự đoán, xây dựng giao diện và đánh giá hiệu quả của hệ thống.

2. Tổng quan về vấn đề nghiên cứu

Hệ khuyến nghị (Recommender Systems) là một lĩnh vực quan trọng trong TTNT, mô phỏng cách con người phân tích thông tin, đánh giá lựa chọn và đưa ra quyết định. Trong ngành tài chính – ngân hàng, hệ khuyến nghị được triển khai nhằm hỗ trợ khách hàng sản phẩm phù hợp với nhu cầu riêng, góp phần nâng cao hiệu quả sử dụng dịch vụ và cải thiện trải nghiệm cũng như mức độ hài lòng của họ.

Về mặt kỹ thuật, ba hướng tiếp cận chính thường được áp dụng trong xây dựng hệ khuyến nghị bao gồm:

- Hệ thống dựa trên luật (Rule – based Systems): hoạt động dựa trên tập quy tắc IF – THEN, dễ diễn giải và phù hợp các nghiệp vụ có tính quy định cao.
- Hệ thống khuyến nghị dựa trên dữ liệu (Recommendation Systems): sử dụng mô hình học máy như Content – based Filtering và Collabrative Filtering nhằm nhận diện sở thích và nhu cầu của người dùng.

- Mô hình khuyến nghị lai (Hybrid): kết hợp các ưu điểm của hai hướng trên nhằm cải thiện chất lượng gợi ý và khả năng đáp ứng nhu cầu cá nhân [6], [8].

Trong bối cảnh ngân hàng hiện đại, việc phát triển một mô hình tích hợp giữa hệ luật và hệ khuyến nghị được xem là hướng tiếp cận hứa hẹn. Hệ thống như vậy không chỉ có năng lực xử lý, đánh giá thông tin của khách hàng, cung cấp gợi ý phù hợp, mà còn đảm bảo khả năng giải thích – yếu tố giữ vai trò trọng yếu trong lĩnh vực tài chính vốn yêu cầu tính minh bạch cao. Đây chính là nội dung nghiên cứu chủ đạo của đề tài.

3. Mục đích nghiên cứu

Mục tiêu tổng quát của đề tài là phát triển một mô hình khuyến nghị nhằm hỗ trợ khách hàng cá nhân trong việc lựa chọn dịch vụ ngân hàng phù hợp nhất với nhu cầu và khả năng tài chính của họ. Các mục tiêu chi tiết bao gồm:

- Nghiên cứu và phân tích hành vi, nhu cầu của khách hàng dựa trên dữ liệu mô phỏng, từ đó xác định các yếu tố quyết định ảnh hưởng đến lựa chọn dịch vụ tài chính.
- Thiết kế mô hình khuyến nghị lai kết hợp giữa hệ luật và các kỹ thuật CBF/CF nhằm hiểu và phân tích dữ liệu đầu vào của khách hàng, đưa ra các gợi ý dịch vụ cá nhân hóa, giải thích rõ ràng lý do của từng khuyến nghị.
- Xây dựng nguyên mẫu hệ thống khuyến nghị dựa trên Python và Flask, đảm bảo khả năng sẵn sàng tích hợp trong môi trường thực tế.
- Đánh giá mô hình thông qua các kịch bản thực nghiệm trên tập dữ liệu synthetic (giả lập), đo lường độ chính xác và tính khả dụng của hệ thống.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: các khách hàng cá nhân có nhu cầu sử dụng dịch vụ ngân hàng và mô hình hệ khuyến nghị thông minh hỗ trợ họ đưa ra quyết định sử dụng dịch vụ phù hợp.

Phạm vi nghiên cứu tập trung vào khuyến nghị nhóm dịch vụ tài chính thiết yếu như: mở tài khoản, gửi tiết kiệm, vay tiêu dùng, thẻ tín dụng, đầu tư phổ thông (gửi góp, chứng chỉ quỹ) và ngân hàng số. Dữ liệu sử dụng trong nghiên cứu được

xây dựng dạng giả lập và hệ thống được phát triển dưới hình thức mô hình thử nghiệm.

5. Phương pháp nghiên cứu

Để bảo đảm hoàn thành các mục tiêu đã xác định, đề án triển khai các phương pháp nghiên cứu chủ đạo sau:

- Phân tích tài liệu: thu thập và hệ thống hóa các tài liệu nghiên cứu trong lĩnh vực hệ khuyến nghị, TTNT (Artificial Intelligence - AI) và những ứng dụng chủ đạo trong ngân hàng.
- Khảo sát và phân tích dữ liệu: Thiết kế bộ dữ liệu giả lập mô tả hành vi và nhu cầu của khách hàng cá nhân, đồng thời tiến hành phân tích nhằm xác định các đặc trưng chủ chốt phục vụ mô hình khuyến nghị.
- Thiết kế và xây dựng môhình: Triển khai mô hình hybrid kết hợp rule – based và recommendation system. Rule – based system sẽ được thiết kế dựa trên tập luật dạng IF – THEN, các mô hình CBF và CF được triển khai bằng thư viện scikit - learn.
- Phát triển hệ thống nguyên mẫu: triển khai một ứng dụng web bằng cách sử dụng Python - Flask nằm kiểm chứng khả năng hoạt động của mô hình khuyến nghị.
- Đánh giá mức độ chính xác và hiệu quả của mô hình: tiến hành kiểm tra mô hình bằng nhiều kịch bản, đo lường bằng các chỉ số như MAE, RMSE, NDCG nhằm phản ánh hiệu suất hoạt động và khả năng gợi ý của mô hình.

NỘI DUNG

CHƯƠNG 1: GIỚI THIỆU VỀ HỆ KHUYẾN NGHỊ

Chương này giới thiệu các khái niệm, cơ sở lý thuyết và những mô hình nền tảng liên quan đến hệ thống khuyến nghị, hành vi của người tiêu dùng tài chính cũng như các ứng dụng của hệ khuyến nghị trong ngành ngân hàng. Những nội dung này đóng vai trò làm cơ sở cho việc thiết kế và phát triển hệ thống ở các chương tiếp theo.

1.1. Tổng quan về hệ khuyến nghị

1.1.1. Khái niệm hệ khuyến nghị

Hệ khuyến nghị (Recommender System) là một lĩnh vực quan trọng thuộc TTNT (Artifical Intelligence) và KPDL (Data mining), được thiết kế để hỗ trợ người dùng trong việc ra quyết định bằng cách cung cấp các đề xuất cá nhân hóa dựa trên hành vi, sở thích và đặc điểm riêng của từng người [3], [4], [5]. Về cơ bản, hệ thống này hoạt động như một mô hình dự đoán (predictive model), có nhiệm vụ xác định mức độ quan tâm của người dùng đối với từng sản phẩm hoặc dịch vụ cụ thể dựa trên dữ liệu lịch sử và bối cảnh sử dụng.

Hệ thống vận hành bằng cách thu thập và phân tích dữ liệu từ nhiều nguồn khác nhau như: hồ sơ người dùng (user profile), lịch sử tương tác (interaction history), chuỗi hành vi (behavior sequence), phản hồi trực tiếp/gián tiếp (explicit/implicit feedback), và các thuộc tính mô tả sản phẩm/dịch vụ [3]. Thông qua các kỹ thuật học máy (machine learning) và mô hình thống kê, hệ khuyến nghị có khả năng học được pattern và mối quan hệ tiềm ẩn giữa người dùng và sản phẩm, từ đó đưa ra dự đoán về mức độ phù hợp của từng dịch vụ đối với người dùng cụ thể [5].

Hiện nay, hệ thống khuyến nghị trở thành thành phần quan trọng trong nhiều nền tảng lớn. Trong lĩnh vực thương mại điện tử, các nền tảng như Amazon và Shopee ứng dụng hệ khuyến nghị để đề xuất sản phẩm phù hợp, từ đó nâng cao tỷ lệ chuyển đổi. Trong ngành giải trí, Netflix và Spotify sử dụng các đề xuất cá nhân hóa để tối ưu hóa trải nghiệm người dùng dựa trên thói quen xem và nghe [4], [5]. Trên các mạng xã hội như Youtube và Tiktok, hệ thống khuyến nghị còn đóng vai trò then chốt trong việc định hướng hành vi tiêu thụ nội dung của người dùng.

Xu hướng gần đây cho thấy hệ khuyến nghị ngày càng được ứng dụng mạnh mẽ trong lĩnh vực tài chính – ngân hàng [8], [12], [13]. Khác với các lĩnh vực phổ thông, dữ liệu tài chính có đặc thù nhạy cảm, đa chiều và phụ thuộc vào nhiều yếu tố như thu nhập, phân khúc khách hàng, mục tiêu tài chính, khẩu vị rủi ro và lịch sử sử dụng dịch vụ. Do đó, hệ khuyến nghị trong ngân hàng thường đòi hỏi sự kết hợp giữa mô hình dữ liệu và cơ chế giải thích (explainability) để đảm bảo tính minh bạch và tuân thủ quy định. Hệ thống có thể đưa ra gợi ý như: lựa chọn thẻ tín dụng phù hợp, gói vay ưu đãi, sản phẩm tiết kiệm sinh lời, hay giải pháp bảo hiểm cá nhân.

Ngoài tác động trực tiếp đến trải nghiệm khách hàng, hệ khuyến nghị còn hỗ trợ các ngân hàng tối ưu chiến lược tiếp thị, tăng tỷ lệ chuyển đổi, nâng cao khả năng giữ chân khách hàng và cải thiện hiệu quả phân phối dịch vụ [13], [14]. Nhờ khả năng dự đoán chính xác nhu cầu, ngân hàng có thể triển khai các chiến dịch cá nhân hóa, đồng thời giảm tải rủi ro trong tư vấn thủ công.

Nhìn chung, hệ thống khuyến nghị đóng vai trò then chốt trong các ngân hàng số hiện đại, hỗ trợ tự động hóa quy trình, nâng cao hiệu quả hoạt động và mang đến trải nghiệm tài chính cá nhân hóa cho từng khách hàng.

1.1.2. Phân loại hệ khuyến nghị

Dựa trên phương pháp tiếp cận và loại dữ liệu được sử dụng, hệ khuyến nghị (Recommendation Systems) có thể được phân loại thành bốn nhóm chính: dựa trên luật (rule – based), lọc theo nội dung (Content – based Filtering), lọc cộng tác (Collaborative Filtering), và mô hình lai (Hybrid Recommendation) [4], [5], [6].

- **Hệ khuyến nghị dựa trên luật (Rule – based Recommendation - RB)**

Hệ thống khuyến nghị dựa trên luật là một trong những phương pháp tiên phong, xuất phát từ lĩnh vực hệ chuyên gia và các hệ thống hỗ trợ ra quyết định. Phương pháp này vận hành dựa trên một tập hợp các luật suy diễn (inference rules) do chuyên gia xây dựng thủ công, phản ánh chính xác tri thức nghiệp vụ của lĩnh vực áp dụng.

Ở dạng cơ bản, mỗi luật được mô tả theo cấu trúc IF – THEN, trong đó:

- Vế IF mô tả điều kiện hoặc tập điều kiện phải thỏa mãn (ví dụ: độ tuổi, nghề nghiệp, thu nhập, nhu cầu).
- Vế THEN biểu thị hành động hoặc khuyến nghị cần đưa ra (ví dụ: gợi ý mở thẻ tín dụng hạng chuẩn, đề xuất gói tiết kiệm phù hợp).

Hệ thống rule – based sử dụng bộ suy diễn (Inference Engine) để so khớp thông tin đầu vào của người dùng với các luật hiện có. Hai chiến lược suy diễn phổ biến gồm:

- Forward chaining (suy diễn tiến): xuất phát từ dữ liệu đầu vào → kích hoạt các luật phù hợp → sinh ra khuyến nghị.
- Backward chaining (suy diễn lùi): xuất phát từ khuyến nghị mục tiêu → kiểm tra xem điều kiện của luật có được thỏa mãn hay không.

Nhờ dựa vào tri thức chuyên gia, phương pháp này có các đặc trưng rõ rệt:

- Tính minh bạch cao: quy trình ra quyết định của hệ thống có thể được truy vết một cách dễ dàng, bởi mỗi khuyến nghị đều có thể giải thích dựa trên những quy tắc và điều kiện được định nghĩa trước.
- Dễ kiểm soát & linh hoạt: phù hợp với các lĩnh vực yêu cầu tuân thủ nghiêm ngặt các quy định nghiệp vụ, chẳng hạn như ngân hàng, bảo hiểm hay y tế.
- Không phụ thuộc dữ liệu: hoạt động tốt ngay cả khi dữ liệu lịch sử khan hiếm hoặc không đầy đủ - điều mà các mô hình học máy thường gặp khó khăn.

Tuy nhiên, hệ khuyến nghị dựa trên luật cũng tồn tại nhiều hạn chế đã được chỉ ra trong các nghiên cứu về hệ chuyên gia:

- Không có khả năng tự học: hệ thống không tự cập nhật theo thay đổi hành vi người dùng hoặc xu hướng thị trường, do tất cả tri thức đều được “mã hóa cứng” (hardcoded).
- Khó mở rộng quy mô: khi số lượng dịch vụ, loại khách hàng và tình huống tăng lên, số lượng luật có thể bùng nổ (rule explosion), gây trở ngại cho việc kiểm soát và cải thiện hệ thống.
- Hạn chế bao phủ tình huống: do tri thức chuyên gia luôn có giới hạn, hệ thống khó bao quát toàn bộ trường hợp thực tế, dẫn đến bỏ sót nhiều gợi ý tiềm năng.

- Dễ lỗi thời: trong môi trường ngân hàng thay đổi nhanh, nếu các luật không được cập nhật kịp thời, chúng sẽ không phản ánh chính xác các sản phẩm hoặc chính sách hiện hành.

Mặc dù tồn tại nhiều hạn chế, các hệ rule – based vẫn được đánh giá là nền tảng quan trọng trong các bài toán khuyến nghị cần yếu tố giải thích, tính tuân thủ chính sách và tính minh bạch – những yêu cầu cốt lõi của ngành tài chính. Do đó, các mô hình khuyến nghị hiện đại thường kết hợp rule – based với các kỹ thuật học máy để tạo thành mô hình lai (Hybrid Recommendation) nhằm tận dụng điểm mạnh của cả hai hướng tiếp cận.

- **Hệ khuyến nghị lọc theo nội dung (Content – based Filtering - CBF)**

Hệ khuyến nghị dựa trên nội dung (CBF) là một trong những phương pháp phổ biến và có nền tảng lý thuyết vững chắc trong lĩnh vực hệ khuyến nghị. Phương pháp này dựa trên giả thiết rằng người dùng có xu hướng quan tâm đến các sản phẩm hoặc dịch vụ có đặc điểm tương tự những mục họ đánh giá cao hoặc từng tương tác trước đó [4], [5]. Khác với Collaborative Filtering, vốn khai thác hành vi của cộng đồng người dùng, CBF tập trung vào việc xây dựng hồ sơ người dùng (user profile) và hồ sơ sản phẩm (item profile) dựa trên các đặc trưng mô tả (features) của từng sản phẩm hoặc dịch vụ.

Trong mô hình này, mỗi sản phẩm được thể hiện thông qua một tập hợp các đặc tính mô tả, chẳng hạn như mục tiêu tài chính, mức độ rủi ro, nhóm khách hàng, kỳ hạn và các đặc điểm dịch vụ. Tương tự, hồ sơ người dùng được hình thành dựa trên lịch sử tương tác, các giao dịch trước đây hoặc thông tin cá nhân từ phía người dùng. Khi hệ thống tiến hành đưa ra đề xuất, các thuật toán sẽ đánh giá mức độ tương đồng giữa hồ sơ người dùng và các sản phẩm có trong hệ thống. Những sản phẩm có mức tương đồng cao hơn sẽ được ưu tiên đề xuất cho người dùng.

Để xây dựng biểu diễn đặc trưng và tính toán mức độ tương đồng, các kỹ thuật quan trọng thường được sử dụng trong CBF bao gồm:

- TF – IDF: phương pháp này đánh giá mức độ quan trọng của từng từ khóa trong mô tả sản phẩm, giúp nhấn mạnh các đặc trưng phân biệt giữa các dịch vụ [4].
- Cosine Similarity: kỹ thuật này đo lường mức độ gần gũi giữa hai vector đặc trưng, được dùng để so sánh hồ sơ người dùng với sản phẩm, cũng như giữa các sản phẩm với nhau [5].
- Vector hóa văn bản (Vectorization): chuyển đổi các mô tả dịch vụ ngân hàng sang dạng số bằng cách sử dụng các phương pháp như Bag – of – Words hoặc Embedding, giúp thuật toán có thể xử lý và tính toán.
- Mô hình học sâu (Deep Learning): Một số phương pháp hiện đại như Word2Vec và BERT được ứng dụng để cải thiện khả năng nắm bắt ngữ nghĩa trong dữ liệu mô tả của sản phẩm.

Ưu điểm

Phương pháp CBF sở hữu nhiều ưu điểm quan trọng:

- Không phụ thuộc vào dữ liệu của người dùng khác, do đó phù hợp với môi trường có số lượng người dùng hạn chế hoặc thiếu dữ liệu tương tác.
- Hoạt động tốt trong tình huống cold – start người dùng (new user), vì hệ thống vẫn có thể khuyến nghị dựa trên mô tả sở thích ban đầu.
- Dễ giải thích, nhờ khả năng thuật lại lý do gợi ý dựa trên đặc trưng tương đồng giữa sản phẩm và hồ sơ người dùng (ví dụ: tương đồng về mục tiêu tài chính, mức rủi ro hoặc phân khúc khách hàng).
- Tính ổn định cao khi dữ liệu ít thay đổi, đặc biệt phù hợp với môi trường ngân hàng – nơi các sản phẩm thường có đặc tính mô tả rõ ràng và ổn định theo thời gian.

Hạn chế:

Mặc dù hiệu quả, CBF cũng tồn tại các nhược điểm đã được nhiều nghiên cứu chỉ ra [6], [7]:

- Nguy cơ hiện tượng chuyên biệt hóa quá mức (overspecialization), hệ thống có thể chỉ đề xuất những sản phẩm tương tự nhau, hạn chế khả năng khám phá và gợi ý các lĩnh vực hoặc sản phẩm mới.
- Chỉ hoạt động hiệu quả khi sản phẩm có mô tả chi tiết và đặc trưng rõ ràng; các dịch vụ ngân hàng trừu tượng hoặc ít thông tin mô tả sẽ gây khó khăn cho mô hình.
- Không tận dụng được xu hướng hoặc hành vi của cộng đồng người dùng, làm cho mô hình hoạt động kém hơn Collaborative Filtering trong bối cảnh cần huấn luyện với dữ liệu quy mô lớn.

Mặc dù vẫn tồn tại một số hạn chế, CBF vẫn giữ vai trò quan trọng trong các hệ khuyến nghị hiện đại. Khi tích hợp vào các mô hình lai (hybrid), CBF không chỉ nâng cao tính minh bạch, ổn định và khả năng giải thích của hệ thống, mà còn cung cấp thông tin về nội dung để bù đắp những hạn chế của Collaborative Filtering.

- **Hệ khuyến nghị lọc cộng tác (Collaborative Filtering - CF)**

Lọc cộng tác (Collaborative Filtering – CF) là một trong những kỹ thuật nền tảng và được ứng dụng rộng rãi trong các hệ thống khuyến nghị hiện đại [3], [6]. Khác với phương pháp lọc theo nội dung, CF không cần đến thông tin mô tả sản phẩm hay hồ sơ chi tiết của người dùng. Thay vào đó, thuật toán dựa trên các mô hình tương tác giữa người dùng và sản phẩm, chẳng hạn như đánh giá, hành vi xem, mua sắm, hoặc các dạng phản hồi khác [4], [5]. Phương pháp này dựa trên giả thiết rằng những người dùng có hành vi tương đồng trong quá khứ sẽ có xu hướng quan tâm đến những sản phẩm tương tự trong tương lai.

Hai hướng tiếp cận chính của CF

- *User – based Collaborative Filtering*

User – based CF xác định những người dùng có lịch sử tương tác tương tự với người dùng hiện tại, được gọi là ‘neighborhood’. Mức độ tương đồng giữa các người

dùng thường được đo bằng các chỉ số như Cosine Similarity, Pearson Correlation hoặc Jaccard Similarity [4], [5].

Sau khi nhóm người dùng tương đồng được xác đinh, hệ thống sẽ đề xuất những sản phẩm mà nhóm này quan tâm nhưng người dùng hiện tại chưa trải nghiệm.

Phương pháp này mô phỏng cách thức con người thường tham khảo ý kiến lẫn nhau, mang tính trực quan và tự nhiên, tuy nhiên khi số lượng người dùng rất lớn, chi phí tính toán sẽ trở nên đáng kể.

- *Item – based Collaborative Filtering*

Item – based CF không tập trung vào người dùng mà phân tích mối quan hệ giữa các sản phẩm. Phương pháp này dựa trên quan sát rằng những sản phẩm thường được nhiều người dùng đánh giá tương tự sẽ có mức độ liên quan cao.

Khi người dùng quan tâm đến một sản phẩm, hệ thống sẽ đề xuất các sản phẩm tương tự dựa trên ma trận đồng xuất hiện (co – occurrence) hoặc đo lường mức độ giống nhau giữa các sản phẩm. Phương pháp này có ưu điểm về tính ổn định, khả năng mở rộng và phù hợp với các nền tảng quy mô lớn như Amazon, nơi item – based CF được chứng minh hoạt động hiệu quả hơn user – based CF [4], [6].

Thuật toán và kỹ thuật trong CF

CF truyền thống hoạt động dựa trên bộ lọc lân cận (neighborhood – based methods), nhưng sau này được mở rộng bởi nhiều kỹ thuật nâng cao như:

- Matrix Factorization (MF), đặc biệt là SVD, SVD++, NMF: giúp phân rã ma trận người dùng – sản phẩm thành các vector đặc trưng ẩn, là bước đột phá trong hệ thống khuyến nghị nhờ khả năng tổng quát hóa mạnh.
- Implicit Feedback Models: sử dụng các tín hiệu gián tiếp như lượt xem, click, thay vì đánh giá tường minh.
- Autoencoders và Deep Learning – based CF: áp dụng mạng nơ – ron để học representation của người dùng và sản phẩm, cho kết quả tốt hơn trong môi trường dữ liệu lớn.

Những kỹ thuật này giúp CF vượt qua nhiều hạn chế của mô hình truyền thống, nhất là vấn đề dữ liệu thưa (sparsity).

Ưu điểm của Collaborative Filtering

- Không phụ thuộc mô tả sản phẩm, phù hợp khi dữ liệu mô tả nghèo nàn hoặc không chuẩn hóa.
- Tận dụng trí tuệ tập thể (collective intelligence) để tạo ra gợi ý tự nhiên và phong phú.
- Có khả năng nhận diện các mối quan hệ tiềm ẩn giữa người dùng và sản phẩm – những yếu tố mà phương pháp lọc theo nội dung (CBF) khó có thể phát hiện.
- Hoạt động tốt khi dữ liệu tương tác nhiều, nguồn sản phẩm đa dạng.

Hạn chế của Collaborative Filtering

Dù mang lại hiệu quả cao, CF vẫn còn các hạn chế đã được nêu trong nhiều nghiên cứu [4], [5], [7]:

- Cold – start problem: Đối với các người dùng hoặc sản phẩm hoàn toàn mới, do thiếu dữ liệu tương tác nên hệ thống hầu như không thể tạo ra các gợi ý đáng tin cậy.
- Data Sparsity (độ thưa dữ liệu): Trong thực tế, ma trận tương tác thường cực kỳ thưa (sparsity lên đến 99%), khiến việc tính toán độ tương đồng khó khăn.
- Scalability (khả năng mở rộng): User – based CF gặp vấn đề khi số lượng người dùng rất lớn, do cần tính toán tương đồng trên quy mô lớn.
- Không thể giải thích rõ ràng: CF khó cung cấp lý do minh bạch cho mỗi gợi ý, khác với rule – based hoặc CBF.

- **Hệ khuyến nghị lai (Hybrid Recommendation)**

Hệ khuyến nghị lai (Hybrid Recommender System) là phương pháp kết hợp đồng thời hai hoặc nhiều kỹ thuật khuyến nghị khác nhau nhằm tận dụng ưu điểm và giảm thiểu hạn chế của từng phương pháp đơn lẻ. Việc tích hợp này giúp hệ thống đạt được hiệu quả dự đoán cao hơn, tăng tính bao phủ (coverage) và cải thiện khả năng cá nhân hóa trong nhiều bối cảnh ứng dụng phức tạp, đặc biệt là trong lĩnh vực tài chính – ngân hàng, nơi yêu cầu tính ổn định và minh bạch rất cao.

Các mô hình lai thường kết hợp giữa một hoặc nhiều kỹ thuật sau: Rule – based (RB), Content – based Filtering (CBF), Collaboratice Filtering (CF), hoặc các phương pháp dựa trên mô hình học máy hiện đại như Matrix Factorization, Deep Learning, Graph – based Recommendation. Sự kết hợp này cho phép hệ thống vừa khai thác tri thức chuyên gia (RB), vừa tận dụng thông tin nội dung sản phẩm (CBF), vừa học theo hành vi người dùng trong cộng đồng (CF), tạo nên cơ chế khuyến nghị toàn diện và linh hoạt.

Theo Burke (2002), có sáu chiến lược kết hợp phổ biến trong các hệ khuyến nghị lai [6]:

- Weighted Hybrid – Kết hợp có trọng số

Các thuật toán khuyến nghị được chạy đồng thời và mỗi thuật toán đóng góp kết quả theo một trọng số xác định. Trọng số có thể cố định hoặc điều chỉnh dựa trên chất lượng dự đoán.

- Switching Hybrid – Chuyển đổi giữa các phương pháp

Hệ thống lựa chọn thuật toán phù hợp nhất tùy vào ngữ cảnh. Ví dụ: dùng rule – based khi người dùng mới (cold – start), và chuyển sang CF khi đủ dữ liệu đánh giá.

- Cascade Hybird – Kết hợp dạng tầng

Một thuật toán đưa ra danh sách gợi ý ban đầu, thuật toán khác refinement và sắp xếp lại, giúp kết quả cuối cùng tinh gọn và chính xác hơn.

- Feature Combination – Kết hợp đặc trưng

Kết quả từ nhiều hệ khuyến nghị được tích hợp và đưa vào một mô hình học máy chung.

- Feature Augmentation – Bổ sung đặc trưng

Kết quả của một mô hình được sử dụng làm đặc trưng (feature) đầu vào tăng cường cho mô hình khác. Ví dụ: embedding từ CF được dùng trong mô hình CBF.

- Meta – level Hybrid – Mô hình cấp cao

Một mô hình tạo ra biểu diễn (model) và mô hình khác sử dụng lại biểu diễn đó để ra quyết định khuyến nghị.

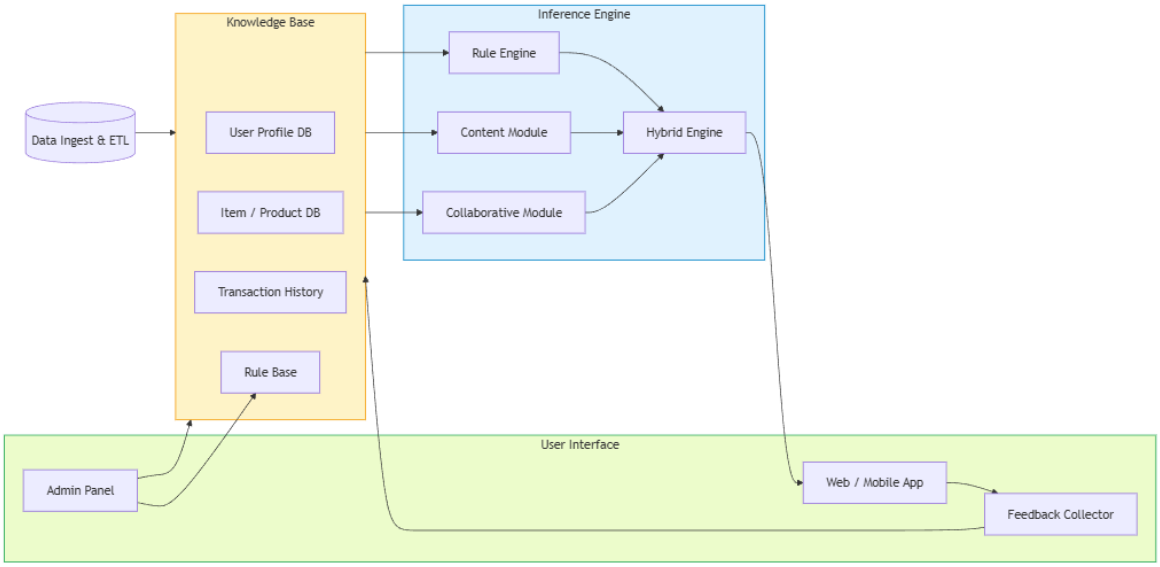
Trong lĩnh vực tài chính – ngân hàng, hệ thống khuyến nghị lai được đánh giá là phù hợp nhất do đặc thù dữ liệu vừa mang tính nghiệp vụ chặt chẽ, vừa bao gồm nhiều thuộc tính hành vi phức tạp. Thực tế, nhiều ngân hàng quốc tế đang triển khai mô hình kết hợp:

- Rule – based + CF: đảm bảo tính minh bạch của hệ luật và khả năng học hành vi người dùng từ cộng đồng.
- CBF + CF: giúp tăng khả năng gợi ý các dịch vụ mới dựa trên đặc trưng nội dung và lịch sử tương tác.
- RB + CBF + CF: tạo thành mô hình lai toàn diện, vừa dễ giải thích, vừa linh hoạt và chính xác.

Ngoài ra, trong môi trường ngân hàng, nơi các khuyến nghị phải đảm bảo tuân thủ quy định, hạn mức, điều kiện sản phẩm và hồ sơ khách hàng, việc tích hợp Rule – based như một lớp kiểm nghiệm (validation layer) giúp tăng độ tin cậy và đảm bảo các gợi ý không vi phạm quy định nghiệp vụ. Nhờ tận dụng thế mạnh của nhiều kỹ thuật khác nhau, hệ khuyến nghị lai ngày càng trở thành yếu tố quan trọng trong các giải pháp tư vấn tài chính, hỗ trợ quá trình chuyển đổi số của ngân hàng và mang lại trải nghiệm tốt hơn cho khách hàng.

1.1.3. Cấu trúc tổng thể của hệ khuyến nghị

Một hệ khuyến nghị hoàn chỉnh thường bao gồm ba thành phần chính: kho tri thức (knowledge base), bộ suy diễn (inference engine) và giao diện người dùng (user interface) [1], [4], [5]. Các thành phần này hoạt động phối hợp để thu thập dữ liệu, xử lý, phân tích và sinh ra các gợi ý phù hợp với từng người dùng.



Hình 1.1: Cấu trúc tổng thể của hệ khuyến nghị

- **Kho tri thức (Knowledge Base)**

Kho tri thức là nền tảng dữ liệu cốt lõi của hệ khuyến nghị, nó lưu trữ và quản lý toàn bộ thông tin liên quan đến người dùng, sản phẩm, tương tác, phản hồi và các luật nghiệp vụ.

Thành phần này có thể bao gồm:

- Cơ sở dữ liệu người dùng (User Profile Database): chứa thông tin định danh, thông tin cá nhân (tuổi, nghề nghiệp, thu nhập, khu vực), hành vi sử dụng và thói quen, mối quan tâm của khách hàng.
- Cơ sở dữ liệu sản phẩm (Item/Product Database): bao gồm mô tả chi tiết sản phẩm, tính chất, mục tiêu, phân khúc khách hàng mục tiêu.
- Dữ liệu lịch sử giao dịch (Transaction & Rating History): ghi lại tương tác giữa người dùng và sản phẩm (giao dịch, đánh giá, tần suất sử dụng).
- Tập luật nghiệp vụ (Rule base): tập hợp các quy tắc "nếu – thì" được phát triển từ kinh nghiệm và hiểu biết thực tiễn của các chuyên gia, phục vụ cho mô hình Rule – based hoặc các điều kiện ràng buộc của hệ thống.

Kho tri thức đóng vai trò nguồn dữ liệu trung tâm, được cập nhật thường xuyên nhằm đảm bảo tính chính xác và khả năng học hỏi của mô hình.

- **Bộ suy diễn (inference Engine)**

Bộ suy diễn được xem là trái tim của hệ khuyến nghị, thực hiện quá trình phân tích, học và đưa ra gợi ý. Thành phần này bao gồm các mô hình học máy và thuật toán được lựa chọn tùy theo loại hệ thống:

  - Rule – based Engine: sử dụng tập luật do chuyên gia định nghĩa để đưa ra các khuyến nghị theo các điều kiện cụ thể.
  - Content – based Module: phân tích đặc trưng nội dung của sản phẩm và hồ sơ người dùng bằng các phương pháp như TF – IDF, Cosine Similarity, hoặc Embedding Models.
  - Collaborative Module: tận dụng dữ liệu tương tác giữa người dùng và sản phẩm được biểu diễn trong ma trận user – item để phát hiện các mối liên hệ, áp dụng Item – based CF hoặc User – based CF để dự đoán sở thích.
  - Hybrid Engine: kết hợp nhiều mô hình (CF, CB, RB) theo các cơ chế như weighted, switching hoặc cascade, tối ưu dựa trên đặc điểm người dùng, giúp khắc phục vấn đề cold – start và tăng độ chính xác.

Bộ suy diễn cũng có thể bao gồm các cơ chế học thích nghi (adaptive learning) giúp mô hình cải thiện theo thời gian thông qua phản hồi của người dùng (feedback loop).

- **Giao diện người dùng (User Interface)**

Giao diện người dùng được xem như kênh liên kết giúp người dùng cuối thao tác và tương tác với hệ thống. Thành phần này có nhiệm vụ hiển thị kết quả gợi ý, thu thập phản hồi và cho phép người dùng tương tác trực tiếp với hệ thống.

Những tiêu chí cơ bản đối với giao diện bao gồm:

  - Trình bày danh sách sản phẩm gợi ý một cách trực quan, có thể sắp xếp theo điểm phù hợp hoặc danh mục sản phẩm.
  - Cung cấp chức năng để người dùng – đặc biệt là người dùng lần đầu – nhập thông tin cá nhân hoặc mục tiêu mà họ mong muốn.
  - Giải thích lý do gợi ý.

- Cung cấp cơ chế thu thập đa phản hồi từ người dùng nhằm nâng cao độ chính xác và hiệu suất của mô hình trong những lần gợi ý tiếp theo.

Trong các hệ thống ngân hàng số, giao diện người dùng thường được triển khai dưới dạng ứng dụng web hoặc ứng dụng di động, triển khai bằng các công nghệ như Flask, ReactJS, hoặc tích hợp tực tiếp trong hệ thống CRM (Customer Relationship Management).

1.2. Tổng quan về dịch vụ ngân hàng

1.2.1. Phân loại dịch vụ ngân hàng cho khách hàng cá nhân

Khi ngành ngân hàng bước vào quá trình chuyển đổi số sâu rộng, các ngân hàng hiện đại đã không ngừng đa dạng hóa danh mục sản phẩm, bao gồm cả dịch vụ truyền thống lẫn các giải pháp tài chính hiện đại được phát triển theo định hướng cá nhân hóa nhằm tối ưu trải nghiệm người dùng. Các sản phẩm hướng đến khách hàng cá nhân có thể được phân loại theo mục đích và nhu cầu sử dụng:

- **Dịch vụ tiết kiệm**

Nhóm sản phẩm tiết kiệm gồm nhiều lựa chọn: tiết kiệm có kỳ hạn, không kỳ hạn, tiết kiệm trực tuyến; các gói gửi góp linh hoạt (định kỳ hoặc không định kỳ) và những sản phẩm đặc thù như tiết kiệm an sinh, học đường, hưu trí hay tích lũy kiều hối, …[16]

Mục tiêu của nhóm này là giúp khách hàng quản lý nguồn tiền nhàn rỗi, sinh lãi an toàn và linh hoạt theo nhu cầu tài chính.

Hệ khuyến nghị có thể gợi ý sản phẩm tiết kiệm tối ưu dựa trên thu nhập, thời hạn mong muốn và mức độ chấp nhận rủi ro.

- **Dịch vụ vay vốn**

Dịch vụ này cung cấp nhiều sản phẩm, điển hình như vay tiêu dùng, vay thế chấp, vay mua nhà, mua xe, vay sản xuất – kinh doanh [16]. Đây là nhóm dịch vụ mang lại nguồn thu chính cho ngân hàng, đòi hỏi việc đánh giá rủi ro và khả năng trả nợ của khách hàng dựa trên hồ sơ tín dụng, lịch sử thu thập và hành vi chi tiêu [8], [13].

Hệ khuyến nghị có thể hỗ trợ xác định sản phẩm vay phù hợp và hạn mức tín dụng tối ưu cho từng cá nhân.

- **Dịch vụ đầu tư**

Bao gồm các dịch vụ: trái phiếu cá nhân, quỹ đầu tư, chứng chỉ quỹ, … [16]. Các ngân hàng ngày càng chú trọng vào việc cung cấp công cụ đầu tư linh hoạt, cung cấp các lựa chọn đầu tư phong phú cho khách hàng. Ứng dụng hệ khuyến nghị giúp phân loại rủi ro, từ đó đề xuất chiến lược đầu tư phù hợp.

- **Dịch vụ thẻ tín dụng**

Dịch vụ này gồm thẻ tín dụng nội địa, quốc tế, thẻ liên kết sinh viên [16], … Nhóm này hướng đến mở rộng khả năng chi tiêu và quản lý dòng tiền cá nhân, đồng thời cung cấp ưu đãi, tích điểm và chương trình khách hàng thân thiết. Hệ khuyến nghị có thể đề xuất loại thẻ phù hợp dựa trên thói quen tiêu dùng, thu nhập và lịch sử sử dụng thẻ của khách hàng.

- **Dịch vụ mở tài khoản**

Bao gồm mở tài khoản cơ bản, tài khoản tiêu chuẩn, trả lương, [16] … Đây là nhóm dịch vụ cơ bản, nhưng cũng là bước đầu trong hành trình trải nghiệm khách hàng, đóng vai trò nền tảng cho các dịch vụ ngân hàng khác.

- **Dịch vụ ngân hàng số**

Bao gồm Mobile Banking, Internet Banking, eKYC, quản lý tài chính cá nhân trực tuyến [16], … Ngân hàng số giúp tự động hóa quy trình giao dịch, cho phép người dùng thực hiện mọi thao tác từ xa với trải nghiệm nhanh chóng, bảo mật và thân thiện.

Các hệ khuyến nghị được tích hợp trực tiếp trong ứng dụng ngân hàng số để phân tích hành vi sử dụng và đưa ra gợi ý phù hợp theo thời gian thực.

- **Dịch vụ thanh toán**

Bao gồm chuyển khoản nhanh 24/7, QR Pay, Ví điện tử, thanh toán hóa đơn tự động, nộp thuế điện tử [16], …Đây là nhóm dịch vụ thiết yếu, phản ứng tần suất giao dịch và hành vi tiêu dùng của khách hàng. Hệ khuyến nghị có thể gợi ý phương thức thanh toán tối ưu, giúp tiết kiệm chi phí và tối ưu trải nghiệm người dùng.

- **Dịch vụ bảo hiểm**

Bao gồm bảo hiểm nhân thọ, bảo hiểm sức khỏe, bảo hiểm tài sản [16], … Ngân hàng thường kết hợp với công ty bảo hiểm để cung cấp giải pháp tài chính toàn diện cho khách hàng. Các mô hình khuyến nghị trong lĩnh vực này giúp xác định đúng nhu cầu bảo vệ và đầu tư, từ đó tăng khả năng tư vấn thành công.

Việc phân nhóm sản phẩm theo mục tiêu sử dụng giúp ngân hàng dễ dàng triển khai chiến lược khuyến nghị cá nhân hóa, từ đó tăng khả năng tiếp cận và mức độ hài lòng của khách hàng.

1.2.2. Phân loại nhu cầu và hành vi người dùng

Mỗi khách hàng cá nhân đều có đặc điểm, mục tiêu và lịch sử sử dụng dịch vụ tài chính khác nhau, chịu ảnh hưởng bởi các yếu tố như thu nhập, độ tuổi, nghề nghiệp, vị trí địa lý và khả năng tiếp cận công nghệ. Các hệ khuyến nghị trong ngân hàng cần xây dựng hồ sơ người dùng (user profile) phản ánh các đặc trưng này để đưa ra gợi ý phù hợp.

Hành vi và nhu cầu tài chính của khách hàng có thể được phân loại theo các tiêu chí:

- Mục tiêu tài chính (Financial Goal): Khách hàng có thể có nhiều mục tiêu tài chính khác nhau, như mở tài khoản hoặc sử dụng ngân hàng số, tiết kiệm và tích lũy tài sản, vay vốn hoặc đầu tư kinh doanh, hoặc tham gia các sản phẩm bảo hiểm và bảo vệ tài chính dài hạn.
- Phân khúc khách hàng (Customer Segment): Các khách hàng có thể được phân loại theo phân khúc như sinh viên hoặc người mới đi làm thường sẽ ưu tiên mở tài khoản, sử dụng ngân hàng số, thẻ tín dụng; nhân viên hoặc người đã đi làm ổn định thường quan tâm đến việc tiết kiệm, đầu tư và vay tiêu dùng; hộ kinh doanh hoặc doanh nhân ưu tiên các gói vay hỗ trợ, quản lý tài chính doanh nghiệp và bảo hiểm tài sản; khách hàng có thu nhập cao (khách hàng VIP) thường chú trọng đầu tư, bảo hiểm cao cấp và dịch vụ tư vấn tài chính cá nhân hóa; nhóm người hưu trí thường quan tâm đến các sản phẩm tiết kiệm dài hạn, bảo hiểm sức khỏe, các dịch vụ quản lý tài chính an toàn, ổn định.

- Hành vi tiêu dùng tài chính (Financial Behavior): Các khách hàng có thể khác nhau về tần suất giao dịch, số lượng sản phẩm tài chính mà họ sử dụng, mức độ phản hồi với các chương trình khuyến mãi, cũng như mức độ sử dụng đa kênh như mobile banking và internet banking.

Đây là bước quan trọng để huấn luyện hệ khuyến nghị, giúp mô hình dự đoán chính xác hơn sản phẩm mà người dùng tiềm năng sẽ quan tâm.

1.2.3. Vai trò của hệ khuyến nghị trong lĩnh vực tài chính – ngân hàng

Áp dụng hệ khuyến nghị trong ngành tài chính – ngân hàng mang lại nhiều lợi ích quan trọng cho cả khách hàng và tổ chức cung cấp dịch vụ. Đối với khách hàng, hệ thống giúp cá nhân hóa trải nghiệm bằng cách đề xuất các sản phẩm phù hợp với nhu cầu, mục tiêu tài chính, mức thu nhập và hành vi giao dịch của từng cá nhân, từ đó nâng cao mức độ thuật tiện và chất lượng tương tác với ngân hàng.

Đối với tổ chức tài chính, hệ khuyến nghị góp phần tối ưu hiệu quả kinh doanh thông qua việc tăng tỷ lệ chuyển đổi sản phẩm, cải thiện hiệu quả tiếp thị và hỗ trợ phân phối dịch vụ đúng đối tượng mục tiêu. Đồng thời, hệ thống giúp ngân hàng xây dựng chiến lược chăm sóc khách hàng dựa trên dữ liệu, từ đó nâng cao mức độ hài lòng, duy trì khách hàng và tăng giá trị vòng đời khách hàng (Customer Lifetime Value).

Bên cạnh đó, hệ khuyến nghị còn đóng vai trò như một công cụ hỗ trợ ra quyết định, giúp bộ phận kinh doanh và tiếp thị phân tích xu hướng sử dụng dịch vụ, đánh giá nhu cầu tiềm năng và tối ưu danh mục sản phẩm. Nhờ vậy, ngân hàng có thể xây dựng các chiến lược sản phẩm,chính sách ưu đãi và chương trình chăm sóc phù hợp hơn, góp phần nâng cao hiệu quả vận hành và năng lực cạnh tranh trên thị trường.

1.3. Mô hình khuyến nghị trong lĩnh vực tài chính – ngân hàng

Sự phát triển của Big Data và các kỹ thuật học máy đã thúc đẩy việc ứng dụng mạnh mẽ các mô hình khuyến nghị trong lĩnh vực tài chính – ngân hàng, nhằm cá nhân hóa dịch vụ, nâng cao trải nghiệm khách hàng và tối ưu hóa hiệu quả kinh doanh. Dữ liệu ngân hàng như hồ sơ cá nhân, thu nhập, hành vi giao dịch, lịch sử sử dụng

dịch vụ và mục tiêu tài chính là nguồn thông tin quan trọng giúp các mô hình khuyến nghị đưa ra kết quả chính xác và tương thích với từng nhóm khách hàng.

Khác với hệ khuyến nghị tổng quát ở các lĩnh vực thương mại điện tử hoặc giải trí, hệ khuyến nghị trong ngân hàng có đặc thù cần đảm bảo tính bảo mật dữ liệu, giải thích rõ ràng, và tuân thủ các quy định nghiệp vụ. Vì vậy, các mô hình thường phải được điều chỉnh để phù hợp với bối cảnh kinh doanh và yêu cầu giám sát của tổ chức tài chính.

- **Mô hình khuyến nghị dựa trên luật (Rule – based Recommendation):**

Trong ngân hàng, rule – based được sử dụng rộng rãi để tư vấn các sản phẩm có điều kiện nghiệp vụ rõ ràng, chẳng hạn như yêu cầu thu nhập tối thiểu, phân khúc khách hàng, độ tuổi hay mục tiêu tài chính. Tập luật được xây dựng dựa trên tri thức chuyên gia, cho phép hệ thống đưa ra các quyết định minh bạch, dễ kiểm soát và dễ giải thích.

Phương pháp này có ưu điểm là sự rõ ràng trong giải thích và khả năng đảm bảo tuân thủ quy định nội bộ. Tuy nhiên, hệ thống khó mở rộng, khó tự động thích ứng khi thay đổi sản phẩm hoặc chính sách.

- **Mô hình lọc theo nội dung (Content – based Filtering):**

Trong bối cảnh ngân hàng, CBF khai thác các thuộc tính của dịch vụ như loại sản phẩm, mục tiêu tài chính, mức độ rủi ro, phân khúc khách hàng hướng tới, thời hạn và lợi ích. Hồ sơ người dùng cũng được mô tả dựa trên thông tin tài chính, mục tiêu và hành vi sử dụng. Việc tính toán mức độ tương đồng giúp hệ thống đề xuất các sản phẩm liên quan, như khách hàng quan tâm đến tiết kiệm có thể được đề xuất các sản phẩm gửi góp hoặc chứng chỉ quỹ an toàn.

CBF phù hợp với khách hàng mới, song vẫn gặp hạn chế khi chỉ đề xuất các sản phẩm tương tự, chưa hỗ trợ khám phá sản phẩm mới.

- **Mô hình lọc cộng tác (Collaborative Filtering):**

Trong ngân hàng, CF khai thác lịch sử tương tác của nhiều khách hàng để tìm ra các mẫu chung trong hành vi tài chính. Phương pháp này giúp xác định các sản

phẩm thường được lựa chọn cùng nhau (ví dụ, vay mua nhà đi kèm bảo hiểm tài sản) hoặc nhóm khách hàng có thói quen sử dụng dịch vụ tương đồng.

CF mang lại khả năng học sở thích tiềm ẩn, đặc biệt hiệu quả với khách hàng có nhiều giao dịch. Tuy vậy, CF gặp hạn chế với người dùng mới và sản phẩm mới – một vấn đề phổ biến trong các dịch vụ tài chính.

- **Mô hình khuyến nghị lai (Hybrid Recommendation Model):**

Để tận dụng đầy đủ ưu điểm của từng mô hình, các ngân hàng thường triển khai mô hình lai kết hợp Rule – based, Content – based và Collaborative Filtering. Cách kết hợp có thể là gán trọng số, chuyển đổi theo điều kiện hoặc tăng cường đặc trưng, tùy theo mức độ dữ liệu và đặc điểm của khách hàng.

Trong khuôn khổ đề tài này, mô hình lai được xây dựng theo hướng Adaptive Hybrid, điều chỉnh trọng số giữa các mô hình thành phần dựa trên mức độ hoạt động của khách hàng: người dùng mới ưu tiên Rule – based và CBF, trong khi khách hàng có lịch sử giao dịch phong phú sẽ được ưu tiên gợi ý dựa trên CF. Cách kết hợp này giúp hệ thống đảm bảo tính chính xác, linh hoạt, phù hợp hơn với thực tế vận hành của ngân hàng.

1.4. Kết luận chương 1

Chương 1 đã tổng hợp và trình bày toàn diện các cơ sở lý thuyết nền tảng liên quan đến hệ khuyến nghị và ứng dụng của chúng trong lĩnh vực tài chính – ngân hàng. Trên cơ sở phân tích khái niệm, vai trò và phạm vi ứng dụng của hệ khuyến nghị, chương này đã làm rõ các phương pháp tiếp cận phổ biến gồm: mô hình dựa trên luật, lọc theo nội dung, lọc cộng tác và mô hình lai. Mỗi phương pháp được mô tả theo nguyên lý hoạt động, ưu điểm, hạn chế và tính phù hợp trong môi trường nghiệp vụ ngân hàng.

Bên cạnh đó, chương đã trình bày đặc trưng danh mục sản phẩm ngân hàng dành cho khách hàng cá nhân, đồng thời phân tích hành vi và nhu cầu người dùng theo các tiêu chí như mục tiêu tài chính, phân khúc khách hàng và hành vi tiêu dùng tài chính. Những phân tích này đóng vai trò quan trọng trong việc xây dựng hồ sơ người dùng

(user profile) – thành phần cốt lõi giúp mô hình khuyến nghị cung cấp các gợi ý chính xác và phù hợp.

Ngoài ra, chương đã giới thiệu cấu trúc tổng thể của một hệ khuyến nghị hoàn chỉnh, gồm kho tri thức, bộ suy diễn và giao diện người dùng, cho thấy cách các thành phần này phối hợp với nhau để tạo ra các gợi ý có tính cá nhân hóa. Qua đó, chương đã khẳng định vai trò thiết yếu của hệ khuyến nghị trong việc nâng cao trải nghiệm khách hàng, tối ưu chiến lược kinh doanh và hỗ trợ hoạt động, tư vấn trong các tổ chức tài chính – ngân hàng hiện đại.

Nhìn chung, Chương 1 đóng vai trò định hướng và thiết lập cơ sở lý thuyết vững chắc, giúp hình thành cơ sở lý luận phục vụ cho quá trình thiết kế, xây dựng và triển khai hệ thống khuyến nghị ở các chương tiếp theo. Các khái niệm, mô hình và phân tích được nêu trong chương này là cơ sở để phát triển mô hình khuyến nghị lai được đề xuất trong đề tài.

CHƯƠNG 2: HỆ KHUYẾN NGHỊ CÁC DỊCH VỤ NGÂN HÀNG

Chương này trình bày quá trình thiết kế và xây dựng hệ thống khuyến nghị dịch vụ ngân hàng, bao gồm phân tích người dùng, mô tả nguồn dữ liệu, xác định yêu cầu kỹ thuật, thiết kế kiến trúc tổng thể và lựa chọn các phương pháp dự đoán phù hợp. Mục tiêu của chương là hình thành nền tảng lý thuyết và kỹ thuật để triển khai mô hình khuyến nghị trong môi trường ứng dụng thực tế.

2.1. Phân tích người dùng

Người dùng cá nhân của ngân hàng là nhóm người dùng đa dạng về độ tuổi, nghề nghiệp, thu nhập và mục tiêu tài chính. Việc phân tích người dùng nhằm xác định đặc điểm, hành vi và nhu cầu của từng nhóm, từ đó phục vụ quá trình cá nhân hóa gợi ý sản phẩm.

- **Đặc điểm khách hàng:**

Các đặc điểm cơ bản của khách hàng cá nhân bao gồm:

Độ tuổi: ảnh hưởng trực tiếp đến nhu cầu tài chính và lựa chọn sản phẩm. Ví dụ, khách hàng trẻ tuổi thường ưu tiên các sản phẩm ngân hàng số tiện lợi, thanh toán nhanh, các dịch vụ liên quan đến giáo dục, du học hoặc khởi nghiệp. Trong khi đó, nhóm khách hàng trung niên và hưu trí quan tâm nhiều hơn đến các sản phẩm bảo toàn vốn, tiết kiệm dài hạn, bảo hiểm, cũng như các dịch vụ tư vấn đầu tư án toàn và chăm sóc sức khỏe tài chính lâu dài.

Giới tính: có thể tác động đến cách khách hàng ra quyết định tài chính, ví dụ, một số nghiên cứu chỉ ra rằng nữ giới có xu hướng ưu tiên quản lý rủi ro và tiết kiệm, trong khi nam giới có xu hướng tìm kiếm các cơ hội đầu tư.

Thu nhập và nghề nghiệp: khung thu nhập và nghề nghiệp quyết định khả năng tiếp cận các sản phẩm tài chính phức tạp hơn. Nhóm khách hàng có thu nhập trung bình thường ưu tiên các sản phẩm tiết kiệm và tín dụng cá nhân, trong khi nhóm thu nhập cao hoặc doanh nhân quan tâm đến bảo hiểm, đầu tư tài chính, quỹ hưu trí hoặc các sản phẩm ngân hàng đầu tư.

Khả năng chấp nhận rủi ro: đặc điểm này ảnh hưởng đến loại sản phẩm khách hàng sẵn sàng tham gia, từ các sản phẩm an toàn như tiết kiệm và bảo hiểm, đến các sản phẩm rủi ro cao hơn như cổ phiếu, trái phiếu doanh nghiệp hoặc quỹ đầu tư.

Việc phân tích chi tiết các đặc điểm này không chỉ giúp ngân hàng hiểu rõ hơn về nhu cầu của khách hàng mà còn là tiền đề để phát triển các cơ chế đề xuất cá nhân hóa trong hệ thống khuyến nghị.

- **Hành vi sử dụng dịch vụ**:

Hành vi của khách hàng là một yếu tố quan trọng để phân tích và dự đoán nhu cầu tài chính cá nhân. Các dữ liệu hành vi bao gồm:

Tần suất giao dịch: xác định mức độ tương tác của khách hàng với các sản phẩm ngân hàng, qua đây nhận diện nhóm khách hàng có mức độ tương tác cao và nhóm ít tương tác.

Sản phẩm khách hàng từng sử dụng: cho thấy xu hướng và sở thích của khách hàng, ví dụ, khách hàng sử dụng nhiều sản phẩm tiết kiệm có thể quan tâm đến các gói bảo hiểm hoặc đầu tư ngắn hạn.

Phản hồi với chương trình khuyến mãi: giúp đánh giá mức độ nhạy cảm của khách hàng với các chiến dịch marketing và ưu đãi, từ đó tối ưu hóa cách triển khai chương trình khuyến mãi cho từng phân khúc.

Mức độ tương tác với các kênh ngân hàng số: phản ánh khả năng tiếp nhận công nghệ và dịch vụ ngân hàng số, ví dụ, khách hàng thường xuyên sử dụng ứng dụng di động hoặc Internet Banking có khả năng tiếp nhận các sản phẩm tài chính số cao hơn.

Những dữ liệu hành vi này được xem là dữ liệu chủ chốt giúp các hệ thống khuyến nghị học và dự đoán các sản phẩm phù hợp, từ đó cải thiện trải nghiệm khách hàng.

- **Phân khúc khách hàng**:

Dựa trên sự kết hợp giữa thông tin cá nhân và hành vi tài chính, khách hàng cá nhân có thể được phân loại thành các phân khúc điển hình, bao gồm:

Sinh viên: ưu tiên các sản phẩm tiết kiệm nhỏ, thẻ thanh toán tiện lợi, các dịch vụ hỗ trợ học tập hoặc chi tiêu hàng ngày.

Người đi làm: quan tâm đến quản lý thu nhập, tiết kiệm dài hạn, đầu tư và bảo hiểm.

Doanh nhân và hộ kinh doanh: chủ yếu quan tâm đến các sản phẩm tín dụng, các giải pháp đầu tư, quản lý dòng tiền và bảo hiểm rủi ro.

Hưu trí: ưu tiên bảo toàn vốn, tiết kiệm dài hạn, các sản phẩm chăm sóc sức khỏe và dịch vụ ngân hàng an toàn.

Mỗi phân khúc khách hàng có mục tiêu tài chính, hành vi tiêu dùng và khả năng chấp nhận rủi ro riêng, do đó việc xác định rõ đặc trưng hành vi và nhu cầu của từng nhóm là cơ sở quan trọng để xây dựng các hệ thống khuyến nghị cá nhân hóa. Khi hệ thống nắm bắt chính xác các thông tin này, khả năng cung cấp gợi ý sản phẩm phù hợp sẽ được tối ưu hóa, từ đó nâng cao hiệu quả tư vấn và sự hài lòng của khách hàng.

2.2. Nguồn dữ liệu và đặc điểm

Dữ liệu là yếu tố nền tảng trong bất kỳ hệ thống khuyến nghị nào, quyết định trực tiếp chất lượng dự đoán và hiệu quả gợi ý sản phẩm. Trong hệ thống nghiên cứu này, dữ liệu được sử dụng là dữ liệu mô phỏng (synthetic data), được xây dựng dựa trên các kịch bản nghiệp vụ thực tế của ngân hàng.

- **Thành phần dữ liệu:**

Nguồn dữ liệu mô phỏng được tổ chức thành ba thành phần chính, tương ứng với các thực thể cơ bản trong ngân hàng.

Hồ sơ khách hàng (Customers):

- Ghi nhận các thuộc tính nhân khẩu học cơ bản như tuổi, giới tính, nghề nghiệp, mức thu nhập và vùng địa lý.
- Bao gồm các thông tin liên quan đến mục tiêu tài chính và khả năng chấp nhận rủi ro, phục vụ việc phân loại phân khúc khách hàng.

- Phân khúc khách hàng được xác định dựa trên sự kết hợp giữa đặc điểm nhân khẩu học và hành vi tiêu dùng, ví dụ: sinh viên, người đi làm, doanh nhân hoặc nhóm hưu trí.

Hồ sơ khách hàng là thành phần quan trọng để hệ thống khuyến nghị hiểu và dự đoán sở thích, nhu cầu tài chính cá nhân.

Danh mục sản phẩm (Products):

- Bao gồm tên sản phẩm, nhóm dịch vụ (tiết kiệm, vay, bảo hiểm, ngân hàng số, đầu tư, …), đặc điểm sản phẩm và nhóm mục tiêu khách hàng.
- Các thông tin này giúp hệ thống khuyến nghị xác định sản phẩm phù hợp với từng phân khúc khách hàng, đồng thời hỗ trợ đánh giá sự phù hợp giữa sản phẩm và sở thích của người dùng.
- Việc phân loại sản phẩm theo nhóm dịch vụ và đặc điểm cụ thể giúp mô hình học được mối quan hệ giữa hành vi khách hàng và các loại sản phẩm.

Lịch sử giao dịch (Transactions):

- Ghi nhận các tương tác của khách hàng với sản phẩm, bao gồm mã khách hàng, mã sản phẩm, điểm đánh giá (rating) và thời gian giao dịch.
- Lịch sử giao dịch phản ánh hành vi thực tế và sở thích của khách hàng đối với từng sản phẩm, từ đó cung cấp dữ liệu huấn luyện quan trọng cho các mô hình collaborative filtering hoặc hybrid.
- Thông qua dữ liệu này, hệ thống có thể phát hiện các mẫu hành vi, dự đoán xu hướng sử dụng sản phẩm, từ đó cung cấp các gợi ý cá nhân hóa.

- **Lưu trữ dữ liệu:**

Các tập dữ liệu được quản lý và lưu trữ trên hệ quản trị cơ sở dữ liệu PostgreSQL, dưới dạng các bảng quan hệ, cho phép:

- Truy xuất linh hoạt thông tin khách hàng, sản phẩm và lịch sử giao dịch.
- Thực hiện các thao tác truy vấn phức tạp, như lọc, phân loại, nhóm và đánh giá các chỉ số thống kê.

- Dễ dàng mở rộng khi triển khai trên nền tảng ứng dụng Flask hoặc các hệ thống backend khác, đảm bảo khả năng tích hợp với các mô hình học máy và khuyến nghị.

Việc chuẩn hóa dữ liệu thành ba bảng chính – customers, products và transactions – tạo ra một cấu trúc dữ liệu chuẩn, giúp hệ thống khuyến nghị huấn luyện, đánh giá và triển khai một cách hiệu quả. Đồng thời, cấu trúc này cũng cho phép mở rộng các bảng bổ sung trong tương lai, ví dụ như bảng ghi nhận phản hồi của khách hàng, lịch sử truy cập digital banking ("*ngân hàng số*") hoặc thông tin ưu đãi, nhằm nâng cao khả năng cá nhân hóa và độ chính xác của hệ thống.

- **Vai trò dữ liệu trong hệ thống khuyến nghị**

Dữ liệu mô phỏng không chỉ giúp mô phỏng các tình huống thực tế trong ngân hàng mà còn là nền tảng để huấn luyện các mô hình khuyến nghị, từ các phương pháp dựa trên nội dung (CBF) đến CF và hybrid; đánh giá chất lượng các gợi ý sản phẩm bằng phương pháp so sánh dự đoán với dữ liệu mô phỏng đã biết; cải thiện trải nghiệm người dùng bằng cách cung cấp gợi ý sản phẩm phù hợp với đặc điểm, hành vi và nhu cầu thực tế của từng nhóm khách hàng.

Nhờ vậy, việc xây dựng cũng như quản lý dữ liệu mô phỏng có vai trò trung tâm trong việc phát triển hệ thống khuyến nghị, đảm bảo tính khả thi, chính xác và hiệu quả trong bối cảnh nghiên cứu mà không vi phạm quyền riêng tư của khách hàng.

2.3. Phân tích yêu cầu hệ thống khuyến nghị

Việc phân tích và xác định yêu cầu hệ thống là một bước quan trọng trong quá trình thiết kế hệ thống khuyến nghị, nhằm đảm bảo rằng mô hình phát triển không chỉ đáp ứng được mục tiêu cá nhân hóa sản phẩm cho từng khách hàng mà còn có độ chính xác cao, hiệu năng ổn định và khả năng tích hợp linh hoạt vào môi trường ứng dụng ngân hàng thực tế. Các yêu cầu hệ thống được chia thành hai nhóm chính: yêu cầu chức năng và yêu cầu phi chức năng.

- **Yêu cầu chức năng:**

Yêu cầu chức năng mô tả các tính năng mà hệ thống phải thực hiện được, bao gồm:

Quản lý thông tin khách hàng

Hệ thống cho phép người dùng nhập thông tin của khách hàng mới, bao gồm các thuộc tính nhân khẩu học, mục tiêu tài chính, khả năng chấp nhận rủi ro, hoặc lựa chọn khách hàng đã có dữ liệu trong cơ sở dữ liệu. Việc này giúp hệ thống có thể xây dựng hồ sơ khách hàng đầy đủ và cá nhân hóa, làm cơ sở cho quá trình sinh gợi ý sản phẩm.

Sinh gợi ý sản phẩm cá nhân hóa

Dựa trên hồ sơ khách hàng và lịch sử giao dịch, hệ thống tự động đề xuất danh sách các sản phẩm phù hợp. Các mô hình khuyến nghị có thể sử dụng nhiều phương pháp khác nhau, từ rule – based, content – based fitlering đến collaborative filtering, hoặc hybrid model kết hợp. Gợi ý này nhằm tối ưu hóa khả năng khách hàng chọn được sản phẩm phù hợp với nhu cầu thực tế, từ đó nâng cao trải nghiệm và sự hài lòng.

Giải thích lý do gợi ý

Hệ thống cần cung cấp thông tin minh bạch về lý do đề xuất sản phẩm, ví dụ "Sản phẩm phù hợp với mục tiêu tiết kiệm của khách hàng" hoặc "Sản phẩm được nhiều khách hàng cùng phân khúc sử dụng". Tính năng này giúp tăng độ tin cậy của khách hàng vào hệ thống, đồng thời hỗ trợ quá trình ra quyết định tài chính của người dùng.

- **Yêu cầu phi chức năng**:

Yêu cầu phi chức năng tập trung vào các tiêu chí chất lượng và hiệu năng của hệ thống, bao gồm:

Độ chính xác

Hệ thống cần đảm bảo rằng các gợi ý đưa ra có độ tin cậy cao, được đánh giá qua các chỉ số học máy tiêu chuẩn như Precision@K, NDCG, MAE, RMSE. Độ chính xác cao không chỉ nâng cao hiệu quả khuyến nghị mà còn tăng khả năng tin tưởng và hài lòng của khách hàng.

Hiệu năng và khả năng phản hồi

Hệ thống cần khai thác và xử lý dữ liệu, sinh gợi ý nhanh chóng, đảm bảo khả năng xử lý trực tuyến (real – time). Điều này đặc biệt quan trọng khi triển khai trong môi trường số, nơi khách hàng kỳ vọng trải nghiệm mượt mà và phản hồi tức thì.

Khả năng mở rộng và tích hợp

Hệ thống phải dễ dàng tích hợp với các ứng dụng ngân hàng hiện có như Internet Banking và Mobile Banking, đồng thời cho phép mở rộng khi thêm khách hàng, sản phẩm hoặc kênh giao dịch mới.

Trải nghiệm người dùng Giao diện cần thân thiện, trực quan, dễ thao tác và hiển thị rõ ràng các gợi ý cùng lý do kèm theo. Một giao diện hiệu quả giúp khách hàng nhận biết giá trị của gợi ý, từ đó nâng cao sự tương tác và mức độ hài lòng với dịch vụ ngân hàng.

Việc xác định rõ phạm vi và tiêu chí đánh giá hệ thống khuyến nghị có ý nghĩa quan trọng trong cả khía cạnh nghiệp vụ và kỹ thuật. Về nghiệp vụ, đảm bảo hệ thống cung cấp các gợi ý phù hợp với mục tiêu tài chính và nhu cầu cá nhân của khách hàng, từ đó tối ưu hiệu quả kinh doanh và nâng cao chất lượng dịch vụ. Về kỹ thuật, hệ thống cần đảm bảo độ chính xác, hiệu năng và khả năng mở rộng, đồng thời dễ tích hợp với các nền tảng ngân hàng số hiện có. Nhờ việc phân tích yêu cầu một cách chi tiết và toàn diện, quá trình thiết kế, phát triển và triển khai hệ thống khuyến nghị có thể diễn ra suôn sẻ, giảm thiểu rủi ro kỹ thuật và tăng khả năng đáp ứng yêu cầu và kỳ vọng của khách hàng.

2.4. Thiết kế kiến trúc hệ thống khuyến nghị

Kiến trúc hệ thống được xây dựng theo mô hình bốn lớp nhằm đảm bảo tính module, dễ mở rộng và thuận lợi cho việc bảo trì trong quá trình vận hành. Mỗi lớp đảm nhận một vai trò riêng trong chuỗi xử lý dữ liệu và sinh gợi ý.

Các thành phần chính:

- *Giao diện người dùng (User Interface):*

Lớp này cung cấp môi trường tương tác trực tiếp với người sử dụng. Giao diện được triển khai bằng Flask, cho phép nhập thông tin khách hàng hoặc truy xuất hồ sơ

có sẵn trong cơ sở dữ liệu. Sau khi hệ thống tạo gợi ý, kết quả được trình bày trực quan trên trình duyệt, kèm theo giải thích ngắn gọn về cơ sở của từng đề xuất.

- *Bộ xử lý dữ liệu (Data Processor):*

Lớp xử lý dữ liệu chịu trách nhiệm chuẩn hóa dữ liệu đầu vào nhằm đảm bảo tính nhất quán trước khi đưa vào mô hình. Các chức năng chính bao gồm:

  - Làm sạch dữ liệu (xử lý giá trị thiếu, định dạng không hợp lệ)
  - Chuẩn hóa và mã hóa thuộc tính dạng số và dạng phân loại
  - Xây dựng các biểu diễn vector phù hợp cho từng thuật toán.

Lớp này đồng thời xử lý truy xuất dữ liệu từ PostgreSQL, chuẩn hóa dữ liệu để phù hợp với các mô hình khuyến nghị.

- *Bộ khuyến nghị (Recommendation Engine):*

Bộ máy khuyến nghị là thành phần trung tâm của hệ thống, tích hợp ba phương pháp tiếp cận:

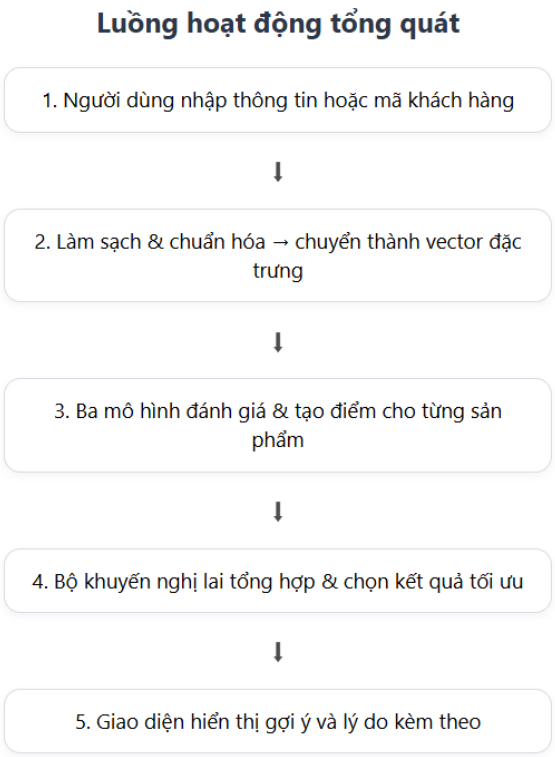
  - Rule - based Filtering: áp dụng tập luật IF – THEN được thiết kế dựa trên đặc điểm hồ sơ khách hàng và thuộc tính sản phẩm.
  - Content – based Filtering: sử dụng TF – IDF và độ tương đồng cosine để đánh giá mức phù hợp giữa mô tả sản phẩm và đặc trưng người dùng.
  - Collaborative Filtering: khai thác lịch sử tương tác và hành vi tương đồng giữa các khách hàng để dự đoán mức độ quan tâm.

Để nâng cao chất lượng dự đoán, hệ thống sử dụng cơ chế Hybrid, kết hợp có trọng số kết quả từ cả ba phương pháp. Việc tổng hợp đa mô hình giúp tận dụng ưu điểm của từng kỹ thuật, đồng thời giảm thiểu các hạn chế như dữ liệu thưa hoặc phụ thuộc nội dung.

- *Kho tri thức và cơ sở dữ liệu (Knowledge Base):*

Lớp này lưu trữ toàn bộ thông tin cốt lõi của hệ thống, bao gồm hồ sơ khách hàng, danh mục sản phẩm, thuộc tính nghiệp vụ, lịch sử tương tác và dữ liệu phản hồi. Kho tri thức đóng vai trò hỗ trợ truy vấn nhanh và phục vụ cho quá trình cập nhật, huấn luyện hoặc tái dự đoán trong những phiên làm việc tiếp theo.

Luồng hoạt động:



Hình 2.1: Luồng hoạt động tổng quát của hệ khuyến nghị

2.5. Phương pháp ước lượng và dự đoán

Dựa trên các khái niệm và mô hình lý thuyết đã trình bày ở Chương 1, phần này mô tả chi tiết cách triển khai phương pháp ước lượng và dự đoán trong hệ thống khuyến nghị dịch vụ ngân hàng, nhằm đảm bảo độ chính xác, khả năng cá nhân hóa và tính minh bạch của kết quả gợi ý. Hệ thống được thiết kế kết hợp nhiều kỹ thuật, khai thác cả tri thức chuyên gia, đặc điểm sản phẩm và hành vi khách hàng.

2.5.1. Phương pháp dựa trên luật (Rule – based System)

Hệ thống dựa trên luật (Rule – based System) mô phỏng quy trình tư vấn truyền thống của chuyên viên ngân hàng thông qua các luật IF – THEN, được xây dựng dựa trên tri thức chuyên gia và đặc điểm hồ sơ khách hàng. Chẳng hạn, “Nếu khách hàng có mục tiêu tiết kiệm và thu nhập dưới 10 triệu đồng thì gợi ý gói tiết kiệm cơ bản”. Ưu điểm của phương pháp này là tính minh bạch, dễ giải thích và đảm bảo tuân thủ quy định nội bộ. Trong hệ thống, các luật được triển khai dưới dạng bảng điều kiện và hành động, cho phép cập nhật linh hoạt khi chính sách hoặc sản phẩm thay đổi.

2.5.2. Phương pháp lọc theo nội dung (Content – based Filtering).

Phương pháp lọc theo nội dung (CBF) tập trung vào việc so sánh đặc trưng của sản phẩm với hồ sơ khách hàng. Các đặc trưng được biểu diễn dưới dạng vector TF – IDF, và mức độ phù hợp được tính toán bằng cosine similarity. Cách tiếp cận này giúp đề xuất các sản phẩm tương tự với những sản phẩm mà khách hàng đã quan tâm hoặc sử dụng, đặc biệt hữu ích đối với người dùng mới, khi hệ thống chưa có dữ liệu cộng tác từ những người dùng khác. Kết quả dự đoán được chuẩn hóa về thang đánh giá 1 - 5, thuận tiện cho việc kết hợp với các phương pháp khác trong mô hình lai.

2.5.3. Phương pháp lọc cộng tác (Collaborative Filtering)

Phương pháp này khai thác hành vi tương tác của nhóm khách hàng để dự đoán sở thích cá nhân. Hệ thống xây dựng ma trận người dùng – sản phẩm và kết hợp user – based CF với item – based CF nhằm tận dụng ưu điểm của cả hai hướng tiếp cận. CF giúp khai thác tri thức từ cộng đồng người dùng, mở rộng phạm vi gợi ý vượt ra ngoài lịch sử sử dụng riêng lẻ và phát hiện các sản phẩm tiềm năng dựa trên thói quen chung. Điểm dự đoán cuối cùng cũng được chuẩn hóa về thang điểm 1 – 5 để đảm bảo khả năng so sánh khi tích hợp với các mô hình khác.

2.5.4. Mô hình lai (Hybrid Recommendation Model)

Mô hình lai được xây dựng nhằm tận dụng ưu điểm của từng phương pháp bao gồm Collaborative Filtering, Content – based Filtering và Rule – based. Cách tiếp cận này giúp khắc phục hạn chế cold – start ở khách hàng mới, khi họ chưa có lịch sử giao dịch, đồng thời tăng tính ổn định và độ tin cậy của kết quả gợi ý.

Trong hệ thống triển khai, mô hình lai được thiết kế theo hướng Adaptive Hybrid, tức trọng số của các thành phần không cố định mà được điều chỉnh dựa trên mức độ hoạt động của khách hàng. Với người dùng mới, hệ thống ưu tiên Content – based và Rule – based để đảm bảo gợi ý vẫn chính xác dù không có thông tin dữ liệu lịch sử của khách hàng. Khi khách hàng có nhiều giao dịch hơn, trọng số của Collaborative Filtering được tăng dần để khai thác hành vi cộng tác của nhóm người dùng tương tự. Điểm dự đoán cuối cùng được tính bằng tổng hợp có trọng số của ba

mô hình và được chuẩn hóa về thang điểm 1 - 5, bảo đảm tính nhất quán và khả năng so sánh giữa các sản phẩm.

Sự kết hợp linh hoạt của các phương pháp trên cho phép hệ thống đáp ứng đa dạng nhu cầu khách hàng, từ người dùng mới đến những người dùng sở hữu lịch sử giao dịch phong phú, đồng thời cân bằng giữa độ chính xác và khả năng giải thích. Nhờ cơ chế adaptive weighting, mô hình có thể cải thiện hiệu quả theo thời gian, tối ưu hóa khả năng cá nhân hóa và đảm bảo các gợi ý phù hợp với đặc điểm tài chính, mục tiêu và hành vi của từng khách hàng, đồng thời tuân thủ các quy định nghiệp vụ ngân hàng và bảo vệ thông tin khách hàng.

2.6. Kết luận chương 2

Chương 2 đã xây dựng nền tảng lý thuyết và kỹ thuật cho hệ thống khuyến nghị dịch vụ ngân hàng thông qua việc phân tích người dùng, mô tả nguồn dữ liệu, xác định yêu cầu hệ thống và thiết kế mô hình kiến trúc tổng thể. Trước hết, việc phân tích đặc điểm và hành vi của từng nhóm khách hàng giúp định hình rõ nhu cầu tài chính, làm nền tảng cho việc cá nhân hóa gợi ý. Bộ dữ liệu mô phỏng được thiết kế đầy đủ và chặt chẽ, phản ánh mối quan hệ giữa khách hàng – sản phẩm – lịch sử giao dịch, tạo tiền đề cho việc huấn luyện và đánh giá mô hình.

Các yêu cầu chức năng và phi chức năng được xác định cụ thể, đảm bảo hệ thống không chỉ đáp ứng mục tiêu nghiệp vụ mà còn phù hợp với khả năng triển khai thực tế trong môi trường ngân hàng số. Kiến trúc hệ thống được thiết kế theo hướng module hóa, tách biệt rõ ràng các lớp xử lý dữ liệu, mô hình dự đoán và giao diện người dùng, giúp đảm bảo khả năng bảo trì, mở rộng và tích hợp.

Bốn phương pháp gợi ý (RB, CBF, CF, Hybrid) được lựa chọn và phân tích dựa trên ưu nhược điểm riêng, tạo thành một khung kỹ thuật toàn diện để tối ưu độ chính xác, khả năng thích ứng và mức độ giải thích của hệ thống. Đặc biệt, mô hình Hybrid cho thấy sự phù hợp cao trong bối cảnh bài toán, nhờ khả năng dung hòa giữa dữ liệu hồ sơ và hành vi cộng tác.

Tổng thể, Chương 2 đã hình thành một thiết kế hoàn chỉnh và có định hướng rõ ràng, tạo nền tảng vững chắc cho việc triển khai thực nghiệm, đánh giá mô hình và xây dựng giao diện ứng dụng được trình bày trong Chương 3.

CHƯƠNG 3: XÂY DỰNG ỨNG DỤNG

Chương này trình bày quá trình hiện thực hóa mô hình khuyến nghị dịch vụ ngân hàng đã được thiết kế ở Chương 2. Các nội dung bao gồm lựa chọn cộng nghệ, xây dựng dữ liệu thử nghiệm, huấn luyện mô hình, phát triển giao diện ứng dụng bằng Flask, và đánh giá hiệu quả của hệ thống.

3.1. Nền tảng và công cụ triển khai

Hệ thống khuyến nghị được triển khai trên nền tảng Python 3.11, một ngôn ngữ được sử dụng rộng rãi trong xử lý dữ liệu và mô hình hóa học máy. Python sở hữu hệ sinh thái thư viện phong phú, hỗ trợ mạnh cho các tác vụ tiền xử lý dữ liệu, xây dựng mô hình, đánh giá và triển khai hệ thống, đặc biệt phù hợp với các ứng dụng khuyến nghị trong lĩnh vực tài chính – ngân hàng.

3.1.1. Các công nghệ và thư viện sử dụng

Bảng dưới đây cung cấp cái nhìn tổng quan về các công cụ, thư viện và mục đích sử dụng trong quá trình xây dựng hệ thống:

Bảng 3.1: Các công nghệ và thư viện sử dụng

Nhóm công cụ	Công cụ / Thư viện	Mục đích sử dụng
Ngôn ngữ lập trình	Python 3.11	Triển khai toàn bộ mô hình và pipeline xử lý
Web framework	Flask	Xây dựng API Backend và giao diện web đơn giản
Cơ sở dữ liệu	PostgreSQL	Quản lý dữ liệu khách hàng, sản phẩm và lịch sử giao dịch
Xử lý và phân tích dữ liệu	Pandas, NumPy	Tiền xử lý, chuẩn hóa dữ liệu và xây dựng bảng đặc trưng
Mô hình học máy	Scikit-learn	Tính TF – IDF, cosine similarity, đánh giá mô hình

Lọc cộng tác	Pandas, NumPy, Scikit-learn	Tính toán điểm dự đoán dựa trên ma trận người dùng – sản phẩm
Đánh giá mô hình	Precision@K, NDCG, MAE, RMSE	Đo lường chất lượng dự đoán và mức độ phù hợp gợi ý
Giao diện và trình bày	HTML, CSS, Bootstrap	Hiển thị danh sách gợi ý và giao diện nhập thông tin khách hàng

3.1.2. Cơ sở lựa chọn công nghệ

Các công nghệ được lựa chọn dựa trên các tiêu chí:

- Tính ổn định và tin cậy: Python và PostgreSQL đều được sử dụng rộng rãi trong hệ thống sản xuất của nhiều tổ chức tài chính.
- Khả năng hỗ trợ phân tích dữ liệu mạnh mẽ: Bộ thư viện Pandas, NumPy, scikit – learn là tiêu chuẩn trong lĩnh vực xử lý dữ liệu, mô hình dự đoán.
- Dễ dàng triển khai và tích hợp: Flask cho phép đóng gói mô hình thành API, thuận tiện trong kết nối đến các hệ thống ngân hàng sẵn có hoặc ứng dụng web.
- Tính mở rộng của mô hình: Nền tảng Python cung cấp khả năng linh hoạt cho việc tích hợp nhiều thuật toán khuyến nghị khác nhau, bao gồm RB, CBF, CF và hybrid.

Việc lựa chọn công nghệ trên nhằm phát triển một hệ khuyến nghị với kiến trúc nhẹ, dễ triển khai, tích hợp thuận tiện vào môi trường ngân hàng số, đồng thời đảm bảo hiệu năng, khả năng mở rộng cho các thuật toán khuyến nghị.

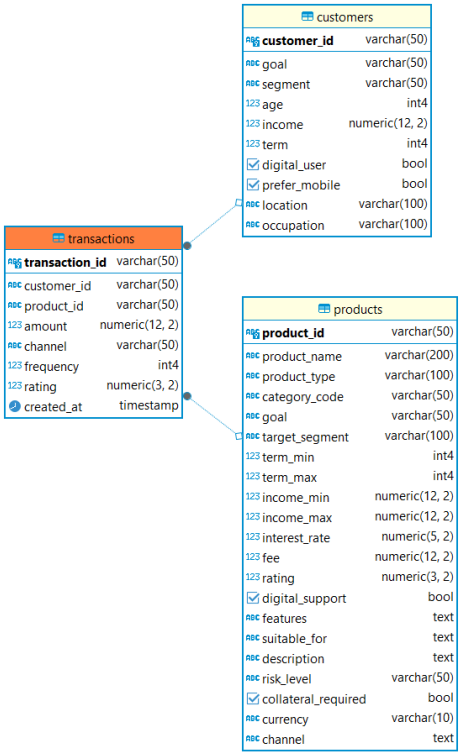
3.2. Xây dựng tập dữ liệu mẫu

Trong thực tế, dữ liệu giao dịch và hành vi khách hàng trong lĩnh vực ngân hàng rất khó tiếp cận do các yêu cầu nghiêm ngặt về bảo mật thông tin và quyền riêng tư. Vì vậy, trong phạm vi đề án, hệ thống khuyến nghị được xây dựng và đánh giá dựa trên một tập dữ liệu mô phỏng (synthetic dataset) thay cho dữ liệu khách hàng thực.

Tập dữ liệu này không được sinh ngẫu nhiên thuần túy, mà được thiết kế và tạo lập dựa trên các quy luật nghiệp vụ ngân hàng bán lẻ, kết hợp với các phân bố xác suất có kiểm soát nhằm phản ánh hợp lý hành vi và nhu cầu của khách hàng cá nhân. Nhờ đó, dữ liệu vẫn đảm bảo tính thực tế, tính đa dạng và tính nhất quán logic giữa các thực thể trong hệ thống.

Bộ dữ liệu gồm ba thành phần chính: hồ sơ khách hàng (Customers), danh mục sản phẩm ngân hàng (Products) và lịch sử giao dịch – đánh giá (Transactions). Ba thành phần này được liên kết chặt chẽ, thể hiện mối quan hệ nhiều – nhiều giữa khách hàng và sản phẩm, đồng thời cung cấp đầy đủ thông tin đầu vào cho các mô hình khuyến nghị được triển khai trong đề án, bao gồm Rule – based, Content – based Filtering và Collaborative Filtering.

Sơ đồ tổng quan cấu trúc dữ liệu được mô tả ở hình dưới, thể hiện ba bảng chính cùng mối quan hệ giữa chúng.



Hình 3.1: Sơ đồ quan hệ giữa các bảng trong tập dữ liệu mô phỏng

3.2.1. Hồ sơ khách hàng (Customers)

Tập dữ liệu khách hàng gồm 1000 hồ sơ, mỗi hồ sơ đại diện cho một khách hàng cá nhân được mô phỏng. Các thuộc tính của khách hàng được sinh theo cơ chế ngẫu nhiên có kiểm soát, trong đó việc sinh dữ liệu không độc lập hoàn toàn mà chịu sự chi phối bởi phân khúc khách hàng (segment) và mục tiêu tài chính (goal). Cách tiếp cận này nhằm đảm bảo dữ liệu phản ánh hợp lý hành vi người dùng và các đặc trưng nghiệp vụ phổ biến trong ngân hàng bán lẻ.

Các trường thông tin chính trong bảng Customers bao gồm:

- customer_id: mã định danh duy nhất cho mỗi khách hàng.
- goal: mục tiêu tài chính chính của khách hàng, bao gồm các nhóm như mở tài khoản, tiết kiệm, vay, thanh toán, ngân hàng số, thẻ tín dụng, bảo hiểm và đầu tư. Mục tiêu được gán theo phân bố xác suất nhằm mô phỏng nhu cầu thực tế của khách hàng cá nhân.
- segment: phân khúc khách hàng, gồm student, low_income, employee, business, vip, retired. Mỗi phân khúc tương ứng với những đặc điểm điển hình về độ tuổi, thu nhập và hành vi sử dụng dịch vụ.
- Age: độ tuổi của khách hàng, được sinh trong khoảng giá trị phù hợp với từng phân khúc (ví dụ: student từ 18-25, retired từ 65 tuổi,…)
- Income: thu nhập trung bình hàng tháng, gán theo mức đặc trưng của từng segment, phản ánh sự chênh lệch thu nhập giữa các nhóm khách hàng.
- Term: thời hạn dự kiến sử dụng sản phẩm (0, 6, 12, 24 hoặc 36 tháng).
- Digital_user: biến nhị phân thể hiện khách hàng có sử dụng dịch vụ ngân hàng số hay không, với tỷ lệ khoảng 70% là người dùng dịch vụ số.
- Prefer_mobile: mức độ ưu tiên sử dụng kênh mobile, chỉ áp dụng cho nhóm digital_user.
- Location: địa điểm sinh sống của khách hàng, được chọn ngẫu nhiên từ danh sách các tỉnh, thành phố nhằm tăng tính đa dạng địa lý.
- Occupation: nghề nghiệp mô phỏng (office_staff, engineer, student, teacher, entrepreneur).

Dữ liệu khách hàng không chỉ được sinh ngẫu nhiên mà còn tuân theo các quy luật nghiệp vụ có điều kiện. Cụ thể, độ tuổi và thu nhập được xác định dựa trên phân khúc khách hàng, giúp tránh các trường hợp phi thực tế như khách hàng sinh viên có thu nhập rất cao hoặc khách hàng đã nghỉ hưu có độ tuổi quá thấp. Bên cạnh đó, các biến hành vi số như digital_user và prefer_mobile được gán với xác suất cao hơn đối với nhóm khách hàng trẻ, phản ánh xu hướng sử dụng kênh số ngày càng phổ biến trong thực tế.

Ngoài ra, mục tiêu tài chính (goal) của khách hàng có mối liên hệ chặt chẽ với phân khúc và mức thu nhập. Nhóm khách hàng thu nhập thấp hoặc sinh viên thường ưu tiên các sản phẩm tiết kiệm, thanh toán cơ bản hoặc ngân hàng số, trong khi các phân khúc có thu nhập cao hơn như business hoặc vip có xu hướng quan tâm nhiều hơn đến các sản phẩm đầu tư, bảo hiểm hoặc tín dụng. Cách sinh dữ liệu này giúp tập Customers duy trì được tính nhất quán logic, đồng thời tạo nền tảng phù hợp cho việc triển khai các mô hình khuyến nghị trong các phần tiếp theo.

3.2.2. Danh mục sản phẩm ngân hàng (Products)

Tập dữ liệu sản phẩm được xây dựng nhằm mô phỏng danh mục dịch vụ đa dạng mà ngân hàng cung cấp, bao gồm các nhóm chính như tài liệu thanh toán, tiết kiệm, tín dụng, bảo hiểm và ngân hàng số. Các trường dữ liệu được thiết kế để phản ánh đầy đủ đặc tính nghiệp vụ cũng như phục vụ cho các mô hình khuyến nghị khác nhau (RB, CBF, CF):

Bảng Products bao gồm các trường:

- Product_id, product_name, product_type, category_code
- Goal: nhóm nhu cầu tài chính mà sản phẩm hướng tới (ví dụ: tiết kiệm, thanh toán, vay, bảo hiểm).
- Target_segment: nhóm khách hàng được sản phẩm hướng tới.
- Term_min, term_max: thời hạn tối thiểu, tối đa áp dụng cho gói sản phẩm.
- Income_min, income_max: khoảng thu nhập yêu cầu đối với khách hàng.
- Interest_rate, fee: các thông số tài chính bao gồm lãi suất và mức phí sử dụng.

- Rating: điểm đánh giá sản phẩm (thang 1 – 5) do hệ thống mô phỏng dựa trên mức độ phù hợp.
- Digital_support: khả năng hỗ trợ giao dịch qua kênh số.
- Features, suitable_for, description: tập thuộc tính, nhóm khách hàng phù hợp và mô tả chi tiết sản phẩm; các trường này phục vụ trực tiếp cho mô hình Content – Based Filtering.
- Risk_level: mức độ rủi ro của sản phẩm, đặc biệt relevant đối với sản phẩm tín dụng hoặc đầu tư.
- Collateral_required: yêu cầu tài sản đảm bảo (nếu có)
- Currency: loại tiền tệ áp dụng.
- Channel: danh sách kênh phân phối (quầy giao dịch, online banking, mobile app, …)

Dữ liệu được sinh ra dựa trên quy luật nghiệp vụ: ví dụ, các sản phẩm tín dụng có mức rủi ro cao hơn và yêu cầu tài sản đảm bảo, trong khi các sản phẩm ngân hàng số luôn hỗ trợ kênh digital. Các trường định lượng như interest_rate, fee, popularity được mô phỏng theo phân phối hợp lý theo từng loại sản phẩm; còn các trường mô tả được tạo theo mẫu để đảm bảo tính thống nhất và hỗ trợ hiệu quả cho việc tính toán TF – IDF trong mô hình Content – Based Filtering.

3.2.3. Lịch sử giao dịch (Transactions)

Tập dữ liệu Transactions được xây dựng nhằm mô phỏng lịch sử tương tác giữa khách hàng và các sản phẩm ngân hàng. Đây là thành phần quan trọng nhất đối với mô hình Collaborative Filtering, bởi nó phản ánh trực tiếp mối quan hệ giữa khách hàng và sản phẩm dưới dạng các cặp (user, item) kèm theo tín hiệu đánh giá. Từ đó, hệ thống có thể xây dựng ma trận người dùng – sản phẩm (user – item matrix) phục vụ cho việc học và dự đoán sở thích.

Mỗi bản ghi trong bảng Transactions đại diện cho một sản phẩm khách hàng đã sử dụng hoặc có tương tác. Các trường dữ liệu bao gồm:

- Customer_id, product_id: khóa ngoại liên kết đến bảng Customers và Products, thể hiện mối quan hệ giữa khách hàng và sản phẩm.

- Amount: giá trị giao dịch hoặc hạn mức sử dụng sản phẩm; giá trị được tạo ngẫu nhiên trong khoảng hợp lý tùy theo loại sản phẩm, phản ánh độ đa dạng trong mức độ sử dụng.
- Channel: kênh giao dịch mà khách hàng sử dụng (quầy giao dịch, Internet Banking, Mobile Banking, …), dữ liệu được lựa chọn dựa trên danh sách kênh mà từng sản phẩm hỗ trợ, đảm bảo tính hợp lệ nghiệp vụ.
- Frequency: tần suất sử dụng sản phẩm (hàng tháng, hàng năm, …), mô phỏng hành vi tiêu dùng thực tế và mức độ gắn bó của khách hàng.
- Rating: điểm đánh giá mức độ hài lòng hoặc mức độ phù hợp của sản phẩm đối với khách hàng, được chuẩn hóa trong thang điểm từ 1 đến 5.

Cơ chế sinh dữ liệu giao dịch

Dữ liệu giao dịch không được sinh ngẫu nhiên thuần túy mà được xây dựng dựa trên ba cơ chế chính, nhằm đảm bảo vừa phản ánh hành vi thực tế vừa đáp ứng yêu cầu huấn luyện mô hình Collaborative Filtering.

(1) Gán sản phẩm theo mục tiêu tài chính

Mỗi khách hàng được gán một nhóm sản phẩm chính phù hợp với mục tiêu tài chính (goal) của họ. Cụ thể, mỗi khách hàng có thể được gán từ 3 đến 6 sản phẩm thuộc cùng nhóm nhu cầu này. Cách tiếp cận này phản ánh thực tế này khách hàng thường tập trung sử dụng các sản phẩm xoay quanh nhu cầu cốt lõi, ví dụ khách hàng có mục tiêu "tiết kiệm" sẽ ưu tiên các gói tiết kiệm khác nhau.

(2) Sinh giao dịch bán chéo (cross-sell)

Bên cạnh sản phẩm chính, hệ thống áp dụng một tập quy tắc bán chéo (cross – sell map) được xây dựng dựa trên chiến lược kinh doanh phổ biến của ngân hàng bán lẻ.

Ví dụ: khách hàng sử dụng sản phẩm tiết kiệm có xu hướng sử dụng thêm ngân hàng số hoặc thẻ tín dụng, khách hàng quan tâm đến vay vốn thường phát sinh nhu cầu bảo hiểm hoặc mở tài khoản thanh toán.

Cơ chế này giúp dữ liệu thể hiện được hiện tượng bán chéo một cách tự nhiên, đồng thời tạo ra mối liên hệ giữa các nhóm sản phẩm khác nhau.

(3) Bổ sung giao dịch để đảm bảo mật độ ma trận

Để đảm bảo ma trận người dùng – sản phẩm không quá thưa, hệ thống sẽ sinh bổ sung các giao dịch ngẫu nhiên có kiểm soát cho các cặp khách hàng – sản phẩm chưa từng tương tác. Việc bổ sung này giúp tăng số lượng giao dịch lên khoảng 20000 bản ghi, đảm bảo dữ liệu phong phú để mô hình Collaborative Filtering học được các mẫu hành vi có ý nghĩa.

Sinh điểm đánh giá (Rating)

Điểm đánh giá rating không được sinh ngẫu nhiên hoàn toàn, mà được xác định dựa trên mức độ phù hợp giữa hồ sơ khách hàng và đặc tính của sản phẩm. Mức độ phù hợp này được tính toán dựa trên các yếu tố như: sự trùng khớp về mục tiêu tài chính (goal), sự phù hợp với phân khúc khách hàng (segment), thu nhập của khách hàng nằm trong khoảng yêu cầu của sản phẩm, khả năng hỗ trợ kênh số đối với khách hàng sử dụng dịch vụ số.

Từ điểm phù hợp này, hệ thống ánh xạ sang một giá trị rating cơ sở, sau đó cộng thêm nhiễu ngẫu nhiên nhỏ, nhằm phản ánh sự khác biệt trong cảm nhận chủ quan của người dùng và tránh việc dữ liệu trở nên quá lý tưởng. Rating cuối cùng được giới hạn trong khoảng từ 1 đến 5 và đóng vai trò là tín hiệu đầu vào chính cho mô hình Collaborative Filtering.

Vai trò của tập dữ liệu Transactions

Nhờ các cơ chế sinh dữ liệu trên, tập Transactions thể hiện được nhiều đặc điểm gần với dữ liệu thực tế: khách hàng có thể tương tác với nhiều nhóm sản phẩm khác nhau, xuất hiện hành vi bán chéo và mức độ đánh giá đa dạng. Ma trận người dùng – sản phẩm có mật độ cao hơn so với dữ liệu thực, giúp mô hình Collaborative Filtering hoạt động hiệu quả ngay cả với quy mô dữ liệu vừa phải.

Tập dữ liệu Transactions giữ vai trò trung tâm trong đề án: cung cấp tín hiệu đánh giá cho mô hình Collaborative Filtering, hỗ trợ kết hợp với dữ liệu nội dung trong mô hình Hybrid, và cho phép đánh giá hiệu quả hệ thống khuyến nghị thông qua các chỉ số Precision@5, NDCG@5, MAE và RMSE. Qua đó, dữ liệu trở thành

cầu nối giữa hành vi người dùng mô phỏng và khả năng học của các mô hình khuyến nghị được triển khai.

3.3. Xử lý và huấn luyện dữ liệu

Quá trình huấn luyện được thực hiện qua bốn bước chính:

3.3.1. Tiền xử lý dữ liệu

Trước khi đưa dữ liệu vào huấn luyện các mô hình khuyến nghị, dữ liệu từ ba bảng Customers, Products và Transactions được tiến hành tiền xử lý nhằm đảm bảo tính nhất quán, giảm nhiễu và phù hợp với yêu cầu của các mô hình Rule – based, Content – based và Collaborative Filtering.

Chuẩn hóa dữ liệu văn bản

Các trường dữ liệu dạng văn bản có liên quan đến việc khớp quy tắc và mô tả nội dung sản phẩm bao gồm goal, segment (trong bảng Customers), goal, target_segment, product_name (trong bảng Products) được chuẩn hóa bằng cách chuyển toàn bộ về chữ thường nhằm tránh sai lệch do khác biệt kiểu chữ, loại bỏ khoảng trắng thừa và các giá trị không hợp lệ, thay thế các giá trị thiếu bằng chuỗi rỗng để đảm bảo quá trình xử lý không bị gián đoạn.

Bước chuẩn hóa này giúp tăng độ chính xác khi so khớp luật trong hệ thống rule – based và khi xây dựng đặc trưng nội dung cho mô hình content – based filtering.

Chuẩn hóa dữ liệu định danh và đánh giá

Các trường định danh customer_id và product_id trong cả ba bảng được ép kiểu thống nhất nhằm đảm bảo khả năng liên kết dữ liệu giữa các bảng.

Trường rating trong bảng Transactions được ép kiểu số, xử lý các giá trị lỗi và được giới hạn trong miền [1, 5], phù hợp với thang đo đánh giá mức độ hài lòng của khách hàng đối với sản phẩm.

Xây dựng ma trận người dùng – sản phẩm (User – Item Matrix)

Từ tập dữ liệu Transactions sau tiền xử lý, hệ thống xây dựng ma trận người dùng – sản phẩm (user – item matrix) để phục vụ cho mô hình Collaborative Filtering.

Cụ thể, mỗi phần tử trong ma trận biểu diễn điểm đánh giá trung bình của một khách hàng đối với một sản phẩm. Trong trường hợp một khách hàng cso nhiều giao

dịch với cùng một sản phẩm, giá trị rating được tổng hợp bằng cách lấy trung bình các lần đánh giá.

Quá trình xây dựng ma trận được thực hiện theo các bước:

(1) Nhóm dữ liệu giao dịch theo cặp (customer_id, product_id) và tính rating trung bình.

(2) Chuyển đổi dữ liệu sang dạng ma trận, trong đó hàng đại diện cho khách hàng và cột đại diện cho sản phẩm.

(3) Các cặp khách hàng – sản phẩm chưa từng tương tác được gán giá trị 0, thể hiện chưa có đánh giá.

(4) Loại bỏ các khách hàng có số lượng tương tác nhỏ hơn 2 nhằm đảm bảo dữ liệu đủ thông tin cho việc học tương đồng trong mô hình lọc cộng tác và hạn chế nhiễu.

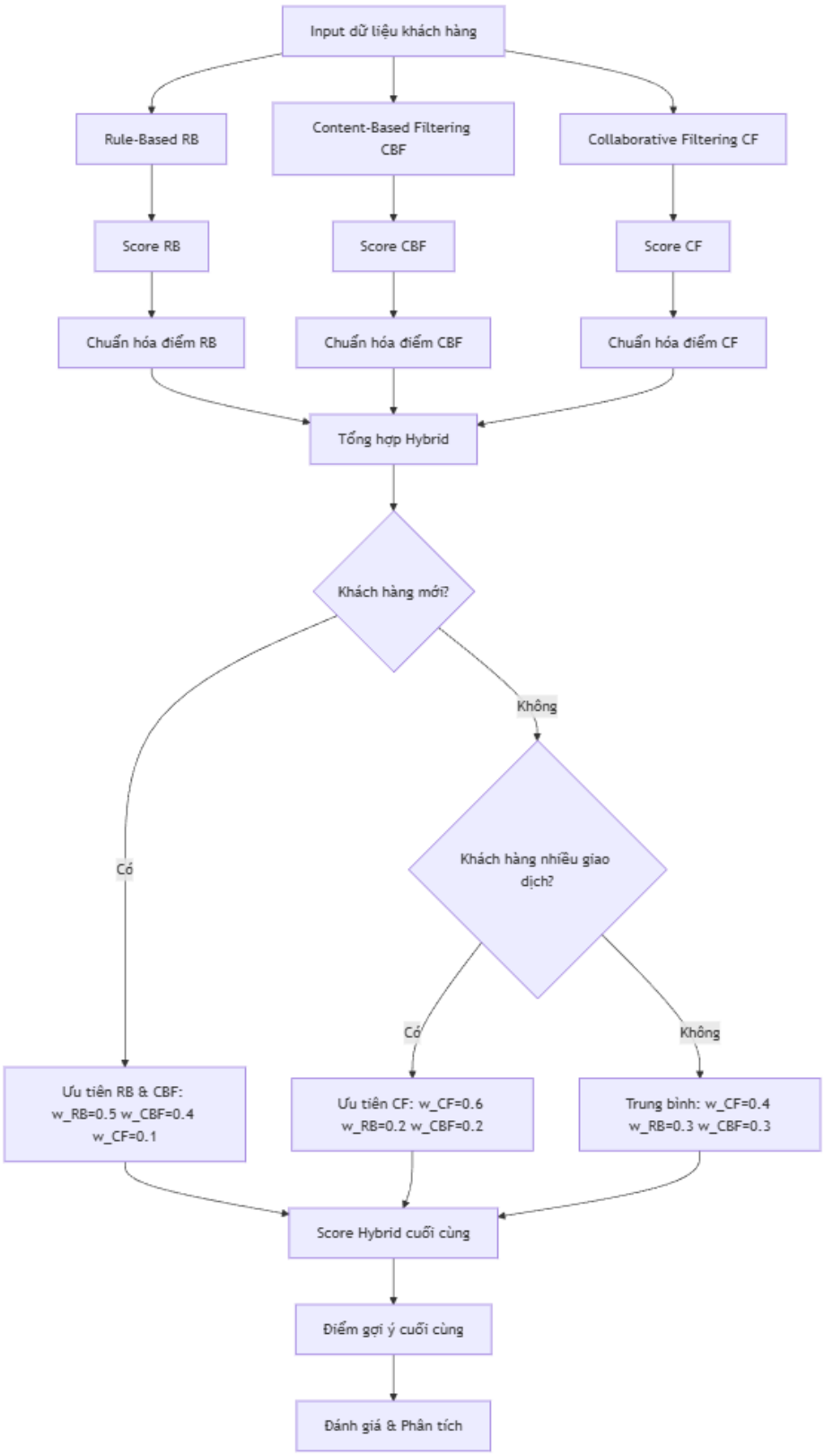
Cách tiếp cận này giúp giảm độ thưa của ma trận và cải thiện hiệu quả học của mô hình Collaborative Filtering trong điều kiện dữ liệu mô phỏng có quy mô vừa phải.

Tách tập huấn luyện và tập kiểm tra

Sau khi xây dựng ma trận user – item, dữ liệu được chia thành hai tập: tập huấn luyện (80%) dùng để huấn luyện mô hình, tập kiểm tra (20%) dùng để đánh giá hiệu năng dự đoán. Việc tách dữ liệu được thực hiện theo từng người dùng. Với mỗi khách hàng, một tập con các sản phẩm đã được đánh giá sẽ được chọn ngẫu nhiên làm tập kiểm tra, các đánh giá tương ứng trong tập huấn luyện được đặt về 0. Cách chia này mô phỏng kịch bản thực tế, trong đó hệ thống cần dự đoán mức độ phù hợp của các sản phẩm mà khách hàng chưa từng sử dụng hoặc chưa từng đánh giá.

3.3.2. Huấn luyện mô hình khuyến nghị

Hệ thống triển khai ba mô hình nền tảng độc lập: rule – based, content – based và collaborative filtering. Sau đó, điểm số từ các mô hình được kết hợp để tạo nên mô hình lai (hybrid recommender).



Hình 3.2: Pipeline huấn luyện mô hình Hybrid

- **Mô hình Rule – based:**

Mô hình rule – based được xây dựng theo tập luật IF – THEN nhằm phản ánh trực tiếp các nguyên tắc nghiệp vụ ngân hàng. Điểm gợi ý cho từng sản phẩm được tính dựa trên ba thành phần:

  - Mức độ phù hợp với mục tiêu (GoalMatch): So sánh giữa mục tiêu tài chính của khách hàng và mục tiêu sản phẩm dựa trên kiểm tra tương đương hoặc bao hàm.
  - Phù hợp phân khúc (SegmentMatch): Đánh giá sự tương đồng giữa phân khúc khách hàng và phân khúc mục tiêu của sản phẩm.
  - Phù hợp thu nhập (IncomeMatch): Kiểm tra việc thu nhập của khách hàng nằm trong khoảng yêu cầu của sản phẩm.
  - Phù hợp nghề nghiệp (OccupationMatch): Xác định mức độ tương thích giữa nghề nghiệp của khách hàng và nhóm nghề nghiệp mà sản phẩm ưu tiên hoặc hướng tới.
  - Phù hợp với kênh số (DigitalMatch): Đánh giá hành vi sử dụng kênh số (mobile/web) của khách hàng so với yêu cầu hoặc nhóm mục tiêu của sản phẩm.
  - Điểm thưởng (Bonus): Cộng điểm bổ sung khi GoalMatch = 1 và SegmentMatch = 1, thể hiện mức độ phù hợp mạnh giữa khách hàng và sản phẩm.

Điểm số cuối cùng được chuẩn hóa về thang [1, 5]:

$$Score_{RB} = 1 + 4 \times (w_{goal} \times GoalMatch + w_{seg} \times SegmentMatch + w_{income} \times IncomeMatch + w_{occupation} \times OccupationMatch + w_{digital} \times DigitalMatch + w_{bonus} \times bonus) \quad (3.1)$$

Mô hình này đồng thời ghi nhận các trường hợp khách hàng và sản phẩm có sự khớp mới mục tiêu mạnh để phục vụ phân tích mở rộng.

- **Mô hình Content – based Filtering:**

Mô hình sử dụng IF – IDF để xây dựng không gian đặc trưng, biểu diễn mô tả sản phẩm. Dữ liệu văn bản kết hợp từ product_name, description, features, và goal tạo thành kho từ vựng cho toàn bộ hệ sản phẩm.

Các bước chính:

- Huấn luyện bộ biến đổi TF – IDF để tạo ma trận biểu diễn đặc trưng cho sản phẩm.
- Mỗi khách hàng được ánh xạ thành một vector hồ sơ dựa trên thông tin goal, segment và occupation.
- Độ tương đồng cosine giữa khách hàng và từng sản phẩm được tính toán.
- Điểm số cuối cùng được tổng hợp từ bốn thành phần:

$$Score_{CB} = 0.4 \times Similarity_{TFIDF} + 0.3 \times GoalMath + 0.2 \times SegmentMatch + 0.1 \times NumbericFit \quad (3.2)$$

Trong đó NumericFit phản ánh mức tương đồng định lượng giữa mức xếp hạng sản phẩm và thu nhập đã chuẩn hóa của khách hàng.

Điểm số được chuẩn hóa về thang [1, 5] để đảm bảo tính nhất quán với các mô hình khác.

- **Mô hình Collaborative Filtering**:

Trong đề án, mô hình Collaborative Filtering (CF) được triển khai nhằm khai thác thông tin từ hành vi và đánh giá của khách hàng đối với các sản phẩm ngân hàng. Do đặc thù dữ liệu mô phỏng có quy mô vừa phải và ma trận người dùng – sản phẩm vẫn tồn tại độ thưa nhất định, đề án áp dụng hai biến thể phổ biến của CF, bao gồm Item – based Collaborative Filtering và User – based Collaborative Filtering, sau đó kết hợp kết quả của hai mô hình để tăng độ ổn định và độ chính xác của dự đoán.

- Item – based CF:

  Ở biến thể item – based CF, trọng tâm là đo lường mức độ tương đồng giữa các sản phẩm dựa trên hành vi đánh giá của người dùng. Ma trận người dùng - sản phẩm được chuẩn hóa theo người dùng bằng cách trừ đi giá trị đánh giá trung bình của mỗi khách hàng. Bước chuẩn hóa này nhằm loại

bỏ thiên lệch cá nhân trong thói quen đánh giá (ví dụ: có khách hàng thường đánh giá cao hơn mặt bằng chung).

Điểm dự đoán của một khách hàng đối với một sản phẩm được tính dựa trên tổng có trọng số giữa mức độ tương đồng của sản phẩm đó với các sản phẩm mà khách hàng đã từng đánh giá và điểm đánh giá tương ứng của khách hàng cho các sản phẩm này.

Cách tiếp cận item – based giúp mô hình hoạt động ổn định ngay cả khi số lượng người dùng lớn và có khả năng khai thác tốt mối quan hệ ngầm giữa các sản phẩm ngân hàng.

- User – based CF:

  Với biến thể User – based CF, mô hình tập trung vào việc tìm kiếm các người dùng có hành vi tương tự nhau. Cụ thể, độ tương đồng giữa các khách hàng được tính toán bằng cosine similarity dựa trên vector đánh giá sản phẩm của từng người dùng. Đối với mỗi khách hàng, hệ thống lựa chọn một tập các người dùng tương tự nhất (nearest neighbors).

  Điểm dự đoán cho một sản phẩm được tính bằng trung bình có trọng số của các điểm đánh giá từ những người dùng tương tự, trong đó trọng số là mức độ tương đồng giữa các người dùng.

  User – based CF cho phép khai thác xu hướng và sở thích chung của các nhóm khách hàng, đặc biệt hữu ích trong việc phát hiện các hành vi tiêu dùng tương đồng chưa được thể hiện rõ ở cấp độ sản phẩm.

- Kết hợp hai biến thể CF:

  Nhằm tận dụng ưu điểm và hạn chế nhược điểm của từng biến thể, điểm dự đoán cuối cùng của mô hình Collaborative Filtering được tính bằng cách kết hợp tuyến tính giữa Item – based CF và user – based CF:

  $$Score_{CF} = 0.6 \times Score_{ItemCF} + 0.4 \times Score_{UserCF} \quad (3.3)$$

  Trong đó, trọng số 0.6 – 0.4 được xác định thông qua thực nghiệm trên tập dữ liệu mô phỏng. Kết quả thử nghiệm cho thấy Item – based CF cho độ ổn định cao hơn trong bối cảnh dữ liệu có số lượng người dùng và sản

phẩm trung bình, trong khi User – based CF bổ sung thông tin về xu hướng hành vi giữa các nhóm khách hàng. Do đó, Item – based CF được ưu tiên trọng số cao hơn.

Điểm dự đoán được cắt ngưỡng trong miền [1,5] nhằm bảo đảm phù hợp với thang đo đánh giá ban đầu và thuận tiện cho việc so sánh, xếp hạng sản phẩm trong giai đoạn khuyến nghị.

- **Mô hình hybrid**

Mô hình lai kết hợp ưu điểm của cả ba phương pháp. Trọng số được điều chỉnh theo mức độ hoạt động của từng khách hàng (số lượng sản phẩm đã tương tác):

- Khách hàng ít tương tác → ưu tiên Rule – based và Content – based.
- Khách hàng trung bình → CF được tăng trọng số.
- Khách hàng hoạt động mạnh → CF được ưu tiên cao nhất do dữ liệu đủ thông tin.

Mô hình lai thu được:

$$Score_{Hybrid} = \omega_{CF} \times CF + \omega_{CB} \times CB + \omega_{RB} \times RB \quad (3.4)$$

Điểm gợi ý cuối cùng được cắt ngưỡng về [1, 5] và dùng cho giai đoạn đánh giá.

Cơ sở xác định trọng số trong mô hình lai

Trong mô hình lai, các trọng số được xác định dựa trên kết hợp giữa cơ sở lý thuyết của hệ thống khuyến nghị và kết quả thực nghiệm trên tập dữ liệu mô phỏng của đề án. Về mặt lý thuyết, Collaborative Filtering cho hiệu quả cao khi dữ liệu lịch sử người dùng đủ lớn, trong khi Rule-Based và Content-Based Filtering hoạt động ổn định hơn trong trường hợp thiếu dữ liệu (bài toán cold-start). Do đó, số lượng sản phẩm mà khách hàng đã từng tương tác được sử dụng làm tiêu chí phản ánh mức độ đầy đủ của dữ liệu hành vi, từ đó điều chỉnh vai trò của từng mô hình thành phần.

Cụ thể, với khách hàng ít tương tác, mô hình ưu tiên Rule-Based và Content-Based; khi số lượng tương tác tăng, trọng số của Collaborative Filtering được tăng dần; và với khách hàng hoạt động mạnh, Collaborative Filtering được ưu tiên cao nhất do khả năng khai thác hiệu quả các mẫu hành vi học được từ dữ liệu.

Các giá trị trọng số cụ thể được tinh chỉnh thông qua thực nghiệm, bằng cách thử nghiệm nhiều cấu hình trọng số khác nhau và lựa chọn cấu hình cho kết quả tốt nhất theo các chỉ số Precision@5, NDCG@5, MAE và RMSE. Cách tiếp cận này đảm bảo mô hình lai vừa có cơ sở khoa học, vừa phù hợp với đặc điểm dữ liệu của đề án.

3.3.3. Đánh giá mô hình

Phương pháp đánh giá

Quá trình đánh giá được thực hiện trên tập kiểm tra (test set) được tách từ ma trận người dùng – sản phẩm theo tỷ lệ 80/20. Đối với mỗi khách hàng, hệ thống sinh danh sách gợi ý Top–5 sản phẩm chưa từng tương tác trong tập huấn luyện. Các mô hình Rule-Based, Content-Based, Collaborative Filtering và Hybrid đều được đánh giá trên cùng một tập dữ liệu và kịch bản dự đoán nhằm đảm bảo tính công bằng trong so sánh. Các điểm dự đoán được chuẩn hóa về miền [1,5] trước khi đưa vào tính toán các chỉ số đánh giá.

Tiêu chí đánh giá

Quá trình đánh giá được thực hiện dựa trên bốn chỉ số thường dùng trong các hệ thống gợi ý, mỗi chỉ số phản ánh một khía cạnh khác nhau của hiệu quả dự đoán.

Cụ thể:

- Precision@5: biểu thị tỷ lệ các sản phẩm liên quan xuất hiện trong năm gợi ý đứng đầu, qua đó phản ánh mức độ chính xác trực tiếp của danh sách đề xuất.
- NDCG@5 (Normalize Discounted Cumulative Gain) đo lường chất lượng xếp hạng bằng cách gán trọng số cao hơn cho các dự đoán đúng xuất hiện ở vị trí đầu.
- MAE (Mean Absolute Error): thể hiện mức sai lệch tuyệt đối trung bình giữa giá trị dự đoán và giá trị thực tế.
- RMSE (Root Mean Square Error): nhấn mạnh các sai số lớn hơn, giúp đánh giá độ ổn định và mức phân tán của sai số dự đoán.

Kết quả đánh giá

Kết quả đánh giá các mô hình được trình bày trong bảng sau:

Bảng 3.2: Kết quả đánh giá mô hình

Mô hình	Precision@5	NDCG@5	MAE	RMSE
Rule - based	0.1296	0.3106	1.0893	1.3590
Content - based	0.1298	0.3065	1.3009	1.5842
Collaborative	0.0362	0.0611	0.8854	1.0479
Hybrid	0.1308	0.3084	0.8641	1.0222

Kết quả thực nghiệm cho thấy Hybrid đạt hiệu năng toàn diện hơn so với các mô hình đơn lẻ. Mặc dù chỉ số Precision@5 và NDCG@5 tương đương với mô hình Rule – based và Content – based, mô hình Hybrid vượt trội đáng kể ở hai chỉ số MAE và RMSE. Điều này cho thấy phương pháp kết hợp vừa duy trì được mức độ chính xác trong xếp hạng, vừa cải thiện chất lượng dự đoán, qua đó tăng tính ổn định của hệ thống gợi ý. Nhìn chung, mô hình Hybrid chứng tỏ mức độ phù hợp cao trong bối cảnh tư vấn sản phẩm tài chính với yêu cầu gợi ý cá nhân hóa.

3.4. Phát triển giao diện ứng dụng

Giao diện của hệ thống được phát triển trên nền tảng Flask, đóng vai trò trung gian giữa mô hình khuyến nghị và người dùng cuối. Ứng dụng hỗ trợ hai nhóm chức năng chính: đề xuất sản phẩm cho khách hàng hiện đại và định hướng lựa chọn cho khách hàng mới.

3.4.1. Gợi ý cho khách hàng hiện hữu

Đối với khách hàng đã có lịch sử sử dụng dịch vụ, hệ thống truy xuất tập các dự đoán sinh ra từ mô hình Hybrid trong quá trình huấn luyện. Trên cơ sở đó, hệ thống thực hiện sắp xếp danh sách sản phẩm theo điểm dự đoán mức độ phù hợp, lọc bỏ các sản phẩm mà khách hàng đã từng đăng ký hoặc sử dụng, lựa chọn top 5 sản phẩm có điểm số cao nhất để hiển thị.

Cách tiếp cận này khai thác thông tin từ cả lịch sử giao dịch của khách hàng và mối tương quan với các khách hàng khác.

3.4.2. Gợi ý cho khách hàng mới

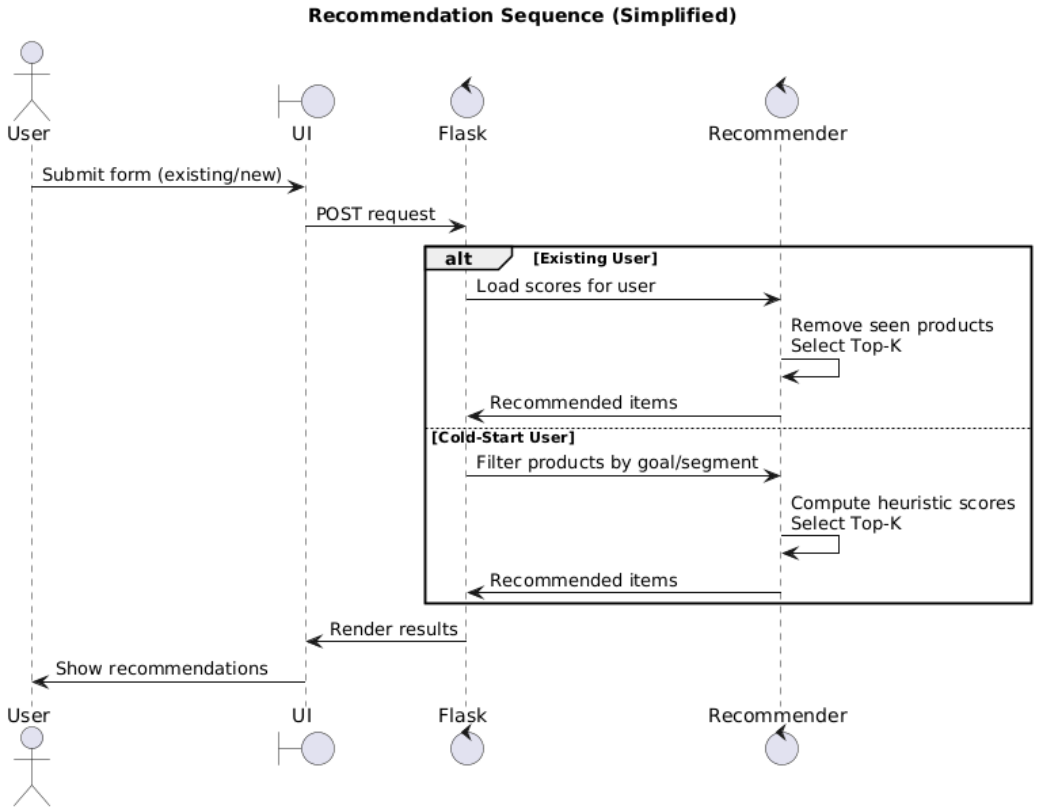
Đối với khách hàng mới, giao diện cho phép nhập một số thông tin cơ bản: độ tuổi, thu nhập, nghề nghề nghiệp, mục tiêu tài chính và mức độ sử dụng kênh số. Sau khi tiếp nhận thông tin đầu vào, hệ thống áp dụng mô hình kết hợp Rule – based (RB), Content – based (CB) để tính toán điểm phù hợp.

Ngoài gợi ý, hệ thống còn cung cấp giải thích chi tiết cho từng sản phẩm, chẳng hạn: "Sản phẩm phù hợp với mục tiêu tài chính", "Các đặc tính sản phẩm tương đồng với nhu cầu của nhóm khách hàng có phân khúc tương tự", … Điều này giúp nâng cao tính minh bạch và tin cậy của hệ thống tư vấn.

3.4.3. Luồng xử lý tổng thể

Quy trình xử lý được tổ chức theo chuỗi các bước:

- Người dùng nhập yêu cầu qua giao diện web.
- Flask tiếp nhận và chuyển dữ liệu đến module xử lý gợi ý.
- Mô hình tính toán danh sách sản phẩm phù hợp.
- Kết quả được trả lại giao diện dưới dạng bảng thông tin, gồm những trường chính: mã sản phẩm (product ID), tên sản phẩm (Product Name), nhóm mục tiêu (Goal), điểm dự đoán (Predicted Score), giải thích lý do gợi ý (Recommendation Reason).



Hình 3.3: Sơ đồ trình tự của hệ thống gợi ý

3.5. Thử nghiệm và đánh giá hệ thống

3.5.1. Kết quả thử nghiệm

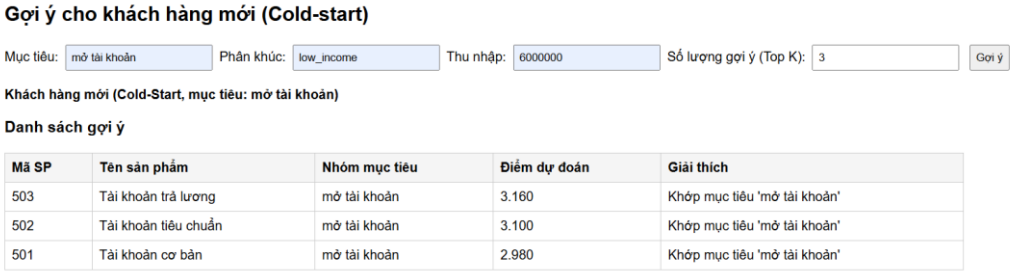
Hệ thống được thử nghiệm trên bộ dữ liệu mô phỏng bao gồm 1000 khách hàng và hơn 30 sản phẩm/ dịch vụ ngân hàng, được sinh theo các quy tắc nghiệp vụ (rule – based) và các phân khúc khách hàng phổ biến. Sau khi huấn luyện các thành phần của mô hình Hybrid và triển khai thử nghiệm toàn bộ quy trình gợi ý, kết quả thu được như sau:

- Precision@5 ≈ 13%: mực độ phù hợp giữa Top – 5 sản phẩm được hệ thống gợi ý và danh mục sản phẩm khách hàng thực sự quan tâm trong dữ liệu mô phỏng.
- Sai số dự đoán đánh giá sản phẩm:
    - MAE ≈ 0.86
    - RMSE ≈ 1.02

Các chỉ số này phản ánh mức độ sai số ở mức trung bình thấp, phù hợp với bản chất dữ liệu synthetic vốn chứa nhiều nhiễu và phân phối phức tạp.

- Thời gian phản hồi trung bình: dưới 0.005s cho mỗi yêu cầu gợi ý trong môi trường cục bộ, đảm bảo khả năng tích hợp vào hệ thống ngân hàng số.

  Ví dụ minh họa:



Hình 3.4: Demo kết quả gợi ý

Các gợi ý đều hợp lý theo logic nghiệp vụ của dữ liệu mô phỏng, cho thấy hệ thống vận hành ổn định và hiệu quả.

3.5.2. Đánh giá và phân tích

Kết quả thử nghiệm chỉ ra rằng mô hình Hybrid cho hiệu quả ổn định trên bộ dữ liệu mô phỏng. Hệ thống có một số ưu điểm và hạn chế như sau:

- Ưu điểm:
  - Mô hình khai thác tốt thông tin hồ sơ khách hàng, mục tiêu tài chính và các quy tắc nghiệp vụ, do đó cải thiện đáng kể vấn đề cold – start so với việc chỉ sử dụng CF.
  - Các luật nghiệp vụ (rule – based) giúp đảm bảo gợi ý luôn nằm trong phạm vi hợp lệ với từng phân khúc, hạn chế gợi ý sai ngữ cảnh.
  - CB và RB vận hành tốt ngay cả khi dữ liệu giao dịch mua sản phẩm chưa phong phú.
- Hạn chế:
  - Thành phần Collaborative Filtering vẫn chưa thể hiện hiệu quả cao do dữ liệu mô phỏng không phản ánh đầy đủ hành vi giao dịch thực tế; lượng tương tác thực chất còn ít và rời rạc.

- Kết quả Precision chưa cao (≈13%) vì dữ liệu synthetic khó mô tả hoàn toàn các mẫu hành vi người dùng thực.

- Đề xuất mở rộng:
  - Bổ sung dữ liệu giao dịch thực hoặc sinh dữ liệu implicit feedback (ví dụ: mức độ quan tâm, lượt xem sản phẩm) để cải thiện ma trận CF.
  - Ứng dụng các mô hình học sâu như Neural Collaborative Filtering (NCF) hay DeepFM khi dữ liệu thật đủ lớn.

3.6. Kết luận chương 3

Chương 3 đã trình bày đầy đủ quy trình xây dựng và triển khai hệ thống khuyến nghị dịch vụ ngân hàng, bao gồm thiết kế bộ dữ liệu mô phỏng, lựa chọn đặc trưng, huấn luyện các mô hình thành phần (RB, CBF, CF) và tích hợp chúng thành mô hình Hybrid. Bên cạnh đó, chương cũng mô tả quá trình hiện thực hóa hệ thống thông qua giao diện ứng dụng xây dựng bằng Flask, cho phép kiểm thử trực tiếp các chức năng gợi ý cho cả khách hàng hiện hữu và khách hàng mới.

Kết quả đánh giá thực nghiệm cho thấy mô hình Hybrid sở hữu độ chính xác và mức độ ổn định cao hơn so với các mô hình đơn lẻ, thể hiện qua các chỉ số đánh giá và các trường hợp thử nghiệm thực tế. Các gợi ý được tạo ra bám sát thông tin hồ sơ, nhu cầu và bối cảnh của từng nhóm khách hàng, qua đó chứng minh tính phù hợp của mô hình trong nghiệp vụ tư vấn sản phẩm tài chính.

Hệ thống nguyên mẫu hoạt động ổn định, đáp ứng yêu cầu kỹ thuật về hiệu năng, thời gian phản hồi và khả năng mở rộng. Các kết quả trên minh chứng hệ thống có tiềm năng triển khai trong thực tế, đồng thời tạo cơ sở cho các hướng mở rộng trong tương lai, như bổ sung dữ liệu thật, ứng dụng mô hình học sâu hoặc tích hợp vào các nền tảng ngân hàng số.

KẾT LUẬN

1. Kết luận.

Đề án đã nghiên cứu, thiết kế và xây dựng thành công một hệ thống khuyến nghị dịch vụ ngân hàng, dựa trên sự kết hợp giữa các kỹ thuật truyền thống và hiện đại trong lĩnh vực Recommender Systems. Hệ thống được thiết kế theo hướng đơn giản, linh hoạt, dễ triển khai, đồng thời đảm bảo tính cá nhân hóa – yếu tố quan trọng hàng đầu trong việc nâng cao trải nghiệm người dùng và tối ưu hóa hiệu quả kinh doanh của ngân hàng.

Trên cơ sở dữ liệu mô phỏng gồm hồ sơ khách hàng, danh mục sản phẩm ngân hàng và lịch sử tương tác, đề án đã áp dụng các kỹ thuật khuyến nghị phổ biến: RB, CBF, CF, và mô hình Hybrid Recommendation. Hệ thống cũng được triển khai dưới dạng ứng dụng web sử dụng Python – Flask, cho phép nhập thông tin khách hàng và hiển thị gợi ý theo thời gian thực, kết hợp giải thích lý do đề xuất.

Thử nghiệm mô hình trên tập dữ liệu giả lập cho thấy Hybrid có kết quả tốt nhất. Điều này cho thấy việc kết hợp ưu điểm của từng mô hình (RB + CB + CF) giúp tăng chất lượng gợi ý và cải thiện khả năng các nhân hóa. Mô hình CF đơn lẻ cho kết quả thấp, phản ánh đúng bản chất: dữ liệu ít, thưa, hành vi người dùng giống nhau chưa đủ rõ ràng để CF hoạt động hiệu quả. Trong khi đó, các thành phần RB và CB bổ sung thông tin ý định và mô tả sản phẩm, giúp cải thiện đáng kể hiệu năng của Hybrid Model.

Thêm vào đó, hệ thống đã:

- Xây dựng được pipeline xử lý dữ liệu, gồm bước làm sạch, chuẩn hóa, mã hóa văn bản (TF – IDF).
- Tạo được cơ chế gợi ý cho cả khách hàng mới (cold – start) bằng cách kết hợp rule – based và content – based.
- Phát triển ứng dụng web hiển thị gợi ý, điểm số dự đoán và lý do khuyến nghị, giúp mô hình có thể vận dụng trong thực tiễn.

Nhìn chung, đề án đã đạt được các mục tiêu đề ra: thiết kế mô hình, xây dựng dữ liệu, triển khai thuật toán, đánh giá hiệu suất và tạo nguyên mẫu hệ thống hoàn chỉnh.

2. Hạn chế.

Mặc dù hệ thống đã vận hành ổn định và đạt độ chính xác tương đối, vẫn còn những hạn chế nhất định:

- Dữ liệu mô phỏng: do không sử dụng dữ liệu ngân hàng thật, mức độ chính xác và tính đại diện còn hạn chế. Hành vi khách hàng mô phỏng không phản ánh đầy đủ các yếu tố thực tế như khả năng tài chính, rủi ro, hành vi chi tiêu.
- Mô hình CF còn đơn giản: Sử dụng phương pháp cosine similarity và biểu diễn ma trận thưa (sparse matrix) ở mức cơ bản, chưa áp dụng các mô hình CF hiện đại như Matrix Factorization hoặc Neural Collaborative Filtering.
- Chưa tối ưu phần giải thích: Hiện tại hệ thống chỉ hiển thị lý do đơn giản, chưa có module giải thích trực quan bằng biểu đồ hoặc mô hình XAI.
- Giao diện demo còn đơn giản, chưa tích hợp xác thực người dùng, chưa mô phỏng được nhiều kịch bản thực tế.

3. Hướng phát triển.

Dựa trên các hạn chế và yêu cầu từ thực tiến ngành tài chính – ngân hàng, hệ thống có thể được mở rộng theo một số hướng sau:

- Mở rộng dữ liệu về tăng tính thực tiễn: Sử dụng dữ liệu thật đã ẩn danh, bổ sung thông tin giao dịch chi tiết, điểm tín dụng và hành vi tài chính để nâng cao độ chính xác của mô hình.
- Nâng cấp mô hình Collaborative Filtering: ứng dụng các thuật toán hiện đại như Matric Factorization, BPR, hoặc Neural Collaborative Filtering nhằm cải thiện chất lượng dự đoán.
- Phát triển mô hình lai thông minh hơn bằng cách tự động hóa việc điều chỉnh trọng số giữa các mô hình thành phần thông qua các kỹ thuật học trọng số hoặc meta – learning.

- Tăng cường khả năng giải thích: cung cấp lý do gợi ý rõ ràng hơn, kết hợp các dạng biểu đồ và mô hình giải thích nhằm nâng cao tính minh bạch và độ tin cậy.
- Triển khai trên môi trường thực tế: đóng gói hệ thống thành API hoặc microservice để tích hợp với ngân hàng số, đồng thời hỗ trợ cập nhật mô hình theo thời gian thực.

4. Kết luận chung

Đề án hoàn thành việc xây dựng một nguyên mẫu hệ thống khuyến nghị dịch vụ ngân hàng dựa trên sự kết hợp giữa kỹ thuật học máy và tri thức nghiệp vụ. Kết quả thử nghiệm cho thấy mô hình Hybrid mang lại hiệu quả tốt nhất, đặc biệt phù hợp với bối cảnh dữ liệu hạn chế và yêu cầu cá nhân hóa cao của ngân hàng.

Hệ thống có thể vận dụng trong thực tiễn khi được mở rộng dữ liệu và nâng cấp thuật toán. Những hướng phát triển đã đề xuất sẽ giúp hoàn thiện mô hình, nâng cao độ chính xác, khả năng giải thích và mở rộng ttinhs ứng dụng trong bối cảnh ngân hàng số hiện đại.

DANH MỤC CÁC TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

[1] Nguyễn Quang Hoan, (2017). *Bài giảng Các hệ thống dựa trên tri thức*, Học viện Công nghệ Bưu chính Viễn thông.

[2] Vũ Hữu Tiệp, (2018). *Machine Learning cơ bản*, 214 – 243

Tài liệu tiếng Anh

[3] Adomavicius G., Tuzhilin A.(2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", IEEE Transactions on Knowledge and Data Engineering.

[4] Ricci F., Rokach L., Shapira B. (Eds.) (2015). *Recommender Systems Handbook*, 2nd ed., Springer, New York

[5] C. Aggarwal, Recommender Systems: The Textbook, Cham, Switzerland: Springer, 2016.

[6] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, p. 331–370, 2002.

[7] J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Systems, vol. 46, p. 109–132, 2013.

[8] D. Zhang and M. Zhou, "A hybrid recommendation system for financial services," International Journal of Information Technology & Decision Making, vol. 19, no. 1, p. 51–78, 2020.

Tài liệu Internet

[9] Experta, "Experta Documentation," 2023. [Online]. Available: https://experta.readthedocs.io/en/latest/introduction.html. Truy cập ngày 02 tháng 07 năm 2025.

[10] LearnOpenCV, "Mastering Recommendation System: A Complete Guide," 2024. [Online]. Available: https://learnopencv.com/recommendation-system/#aioseo-recommendation-system-key-takeaways. Truy cập ngày 05 tháng 07 năm 2025.

[11] SuNT, "Giới thiệu các phương pháp truyền thống xây dựng Recommender System," 2021. [Online]. Available: https://tiensu.github.io/blog/72_recommender_system_1/. Truy cập ngày 05 tháng 07 năm 2025.

[12] M. Fan, "Bank Product Recommendation System," 2023. Survey and Experimen [Online]. Available: https://www.kaggle.com/code/mikefan/bank-product-recommendation-system. Truy cập ngày 110 tháng 07 năm 2025.

[13] Y. Systems, "Recommender Systems for Banking and Financial Services: How AI is Transforming the Way We Manage Our Finances," 2023. [Online]. Available: https://yellow.systems/blog/recommender-systems-for-banking-and-financial-services. Truy cập ngày 01 tháng 07 năm 2025.

[14] KANINI, "Banking Recommendation System: Use Cases," 2023. [Online]. Available: https://kanini.com/blog/banking-recommendation-system-use-cases/. Truy cập ngày 01 tháng 07 năm 2025.

[15] Scaler, "Rule-Based System in AI," 2024. [Online]. Available: https://www.scaler.com/topics/artificial-intelligence-tutorial/rule-based-system-in-ai/. Truy cập ngày 13 tháng 07 năm 2025.

[16] https://www.agribank.com.vn/vn/ca-nhan. Truy cập ngày 01 tháng 07 năm 2025

BẢN CAM ĐOAN

Tôi cam đoan đã thực hiện việc kiểm tra mức độ tương đồng nội dung đề án/luận văn/luận án qua phần mềm Turnitin một cách trung thực và đạt kết quả mức độ tương đồng 17 % toàn bộ nội dung đề án/luận văn/luận án. Bản đề án/luận văn/luận án kiểm tra qua phần mềm là bản cứng đề án/luận văn/luận án đã nộp để bảo vệ trước hội đồng. Nếu sai tôi xin chịu các hình thức kỷ luật theo quy định hiện hành của Học viện.

Hà Nội, ngày tháng 02 năm 2026

HỌC VIÊN CAO HỌC/NCS

(Ký và ghi rõ họ tên**)**

Dương Bảo Anh

17% Tính tương đồng nói chung

Tổng cộng của tất cả các kết quả trùng khớp, bao gồm cả các nguồn trùng lặp, cho mỗi c…

Đã lọc khỏi Báo cáo

- Mục lục tham khảo
- Văn bản được trích dẫn
- Kết quả trùng khớp nhỏ (ít hơn 8 từ)

Nguồn hàng đầu

14% Nguồn Internet
4% Ấn bản
7% Bài tập được nộp (bài của học sinh)

Học viên **Người hướng dẫn khoa học**