# LỜI CẢM ƠN

Để có được một bài báo cáo bổ ích, hiệu quả, hoàn chỉnh và đạt kết quả tốt nhất, trước hết nhóm em xin gửi tới toàn thể các thầy cô khoa Công nghệ thông tin lời chúc sức khỏe, lời chào trân trọng và lời cảm ơn sâu sắc nhất. Đặc biệt, với sự quan tâm, dạy dỗ, chỉ bảo tận tình và chu đáo của thầy Trần Hùng Cường, cùng với sự giúp đỡ nhiệt tình từ các bạn trong nhóm và bạn bè, đến nay nhóm em đã có thể hoàn thành bài báo cáo với đề tài: "Xây dựng hệ thống gợi ý cho người tiêu dùng dựa trên mô hình học sâu".

Nhóm 11 xin bày tỏ lòng biết ơn chân thành và sâu sắc nhất đến thầy Trần Hùng Cường, giảng viên học phần “Kinh doanh thông minh”. Thầy không chỉ truyền đạt cho chúng em những kiến thức chuyên môn vững chắc về kinh doanh thông minh, đặc biệt là về những mô hình học máy và học sâu, các kỹ năng phân tích dữ liệu, lập trình mà còn dành thời gian quý báu để định hướng, vạch kế hoạch chi tiết và hướng dẫn Nhóm 11 hoàn thành bài báo cáo một cách tốt nhất trong suốt thời gian qua. Hơn thế nữa, thông qua quá trình thực hiện đề tài, chúng em đã học được thêm nhiều kiến thức mới, củng cố kiến thức đã học và phát triển thêm một số kỹ năng mềm quan trọng như: kỹ năng làm việc nhóm hiệu quả, kỹ năng quản lý thời gian hợp lý, kỹ năng nghiên cứu và phân tích vấn đề một cách sâu sắc. Chúng em vô cùng cảm ơn thầy đã tận tình chỉ bảo, động viên và tạo điều kiện tốt nhất cho nhóm trong suốt quá trình học tập và thực hiện đề tài.

Các thử nghiệm với mô hình học sâu đã mang lại kết quả khả quan, đồng thời giúp nhóm hiểu rõ hơn về các thuật toán, cũng như cách tối ưu hóa chúng trong thực tế. Tuy nhiên, do giới hạn về thời gian và nguồn lực, hệ thống vẫn còn một số điểm cần hoàn thiện, đặc biệt là trong việc mở rộng dữ liệu và cải thiện độ chính xác của các dự đoán. Nhóm chúng em rất mong nhận được những đóng góp, góp ý từ thầy để có thể cải thiện và hoàn thiện báo cáo. Trong tương lai, nhóm hy vọng có thể tiếp tục nghiên cứu và cải thiện mô hình để hệ thống trở nên mạnh mẽ hơn và ứng dụng được rộng rãi trong nhiều lĩnh vực khác nhau.

Chúng em xin chân thành cảm ơn!

# MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc186816076)

[MỤC LỤC 1](#_Toc186816077)

[DANH SÁCH BẢNG BIỂU 3](#_Toc186816078)

[DANH SÁCH HÌNH ẢNH 3](#_Toc186816079)

[LỜI MỞ ĐẦU 4](#_Toc186816080)

[1. Lý do chọn đề tài 4](#_Toc186816081)

[2. Mục tiêu nghiên cứu 4](#_Toc186816082)

[3. Đối tượng và phạm vi nghiên cứu 5](#_Toc186816083)

[4. Phương pháp nghiên cứu 5](#_Toc186816084)

[5. Kết cấu đề tài 5](#_Toc186816085)

[CHƯƠNG 1. TỔNG QUAN KINH DOANH THÔNG MINH 6](#_Toc186816086)

[1.1. Khái niệm kinh doanh thông minh 6](#_Toc186816087)

[1.1.1. Định nghĩa 6](#_Toc186816088)

[1.1.2. Đặc điểm chính 6](#_Toc186816089)

[1.1.3. Một số công nghệ hỗ trợ BI 6](#_Toc186816090)

[1.1.4. Lợi ích 7](#_Toc186816091)

[1.2. Một số phương pháp và ứng dụng trong kinh doanh thông minh 8](#_Toc186816092)

[1.2.1. Phân tích dữ liệu lớn 8](#_Toc186816093)

[1.2.2. Trí tuệ nhân tạo 8](#_Toc186816094)

[1.2.3. Internet vạn vật 9](#_Toc186816095)

[1.2.4. Điện toán đám mây 9](#_Toc186816096)

[1.2.5. Ứng dụng 9](#_Toc186816097)

[1.3. Những thách thức trong kinh doanh thông minh 10](#_Toc186816098)

[1.4. Các bước cơ bản trong kinh doanh thông minh 11](#_Toc186816099)

[CHƯƠNG 2. HỌC SÂU VÀ HỆ THỐNG GỢI Ý 13](#_Toc186816100)

[2.1. Mạng nơ-ron (Neural network) 13](#_Toc186816101)

[2.1.1. Giới thiệu về học sâu 13](#_Toc186816102)

[2.1.2. Mạng nơ-ron truyền thẳng 16](#_Toc186816103)

[2.1.3. Thuật toán lan truyền ngược (Backpropagation) 21](#_Toc186816104)

[2.2. Hệ thống gợi ý 25](#_Toc186816105)

[2.2.1. Giới thiệu về hệ thống gợi ý 25](#_Toc186816106)

[2.2.2. Các phương pháp tiếp cận khi xây dựng hệ thống gợi ý 28](#_Toc186816107)

[2.2.3. Neural Collaborative Filtering 32](#_Toc186816108)

[2.2.4. Phương pháp và chỉ số đánh giá hệ thống gợi ý 39](#_Toc186816109)

[CHƯƠNG 3. ỨNG DỤNG HỌC XÂY DỰNG HỆ THỐNG GỢI Ý SẢN PHẨM CHO NGƯỜI TIÊU DÙNG 42](#_Toc186816110)

[3.1. Tổng quan về hành vi người tiêu dùng 42](#_Toc186816111)

[3.1.1. Khái niệm 42](#_Toc186816112)

[3.1.2. Vai trò nghiên cứu hành vi người tiêu dùng 42](#_Toc186816113)

[3.2. Tập hợp dữ liệu 43](#_Toc186816114)

[3.3. Phân tích khai phá dữ liệu: 44](#_Toc186816115)

[3.3.1. Xử lý dữ liệu 44](#_Toc186816116)

[3.3.2. Phân tích dữ liệu 47](#_Toc186816117)

[3.4. Huấn luyện mô hình học sâu 53](#_Toc186816118)

[3.5. Đánh giá và so sánh mô hình 55](#_Toc186816119)

[KẾT LUẬN 59](#_Toc186816120)

# DANH SÁCH BẢNG BIỂU

[Bảng 3‑1. Mô hình hành vi người tiêu dùng 42](#_Toc186816121)

[Bảng 3‑2. Bảng thống kê độ thưa của bộ dữ liệu 54](#_Toc186816122)

# DANH SÁCH HÌNH ẢNH

[Hình 2‑1. Hàm Sigmoid 20](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816152)

[Hình 2‑2. Hàm ReLU 20](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816153)

[Hình 2‑3. Hàm Tanh 20](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816154)

[Hình 2‑4. Hàm Leaky ReLU 21](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816155)

[Hình 2‑5. Ma trận tương tác User - Item 27](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816156)

[Hình 2‑6. Quy trình cơ bản của hệ thống lọc cộng tác 29](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816157)

[Hình 2‑7. Hạn chế của MF 33](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816158)

[Hình 2‑8. Kiến trúc NCF 34](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816159)

[Hình 2‑9. Kiến trúc GMF 37](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816160)

[Hình 2‑10. Kiến trúc NeuMF 39](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816161)

[Hình 3‑1. 5 dòng ngẫu nhiên của bộ dữ liệu 44](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816162)

[Hình 3‑2. Tóm tắt bộ dữ liệu 44](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816163)

[Hình 3‑3. 5 dòng đầu tiên có giá trị NaN trong cột Customer ID 45](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816164)

[Hình 3‑4. 5 dòng đầu tiên có giá trị NaN trong cột Description 45](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816165)

[Hình 3‑5. Biểu đồ số lượng dòng trùng lặp hàng tháng 46](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816166)

[Hình 3‑6. 5 dòng trungg lặp đầu tiên 46](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816167)

[Hình 3‑7. Biểu đồ top 10 sản phẩm bán chạy nhất 47](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816168)

[Hình 3‑8. Biểu đồ top 10 sản phẩm bị trả lại nhiều nhất 48](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816169)

[Hình 3‑9. Biểu đồ top 10 sản phẩm có doanh thu cao nhất 48](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816170)

[Hình 3‑10. Biểu đồ cột top 10 quốc gia chiếm doanh thu cao nhất và thấp nhất 49](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816171)

[Hình 3‑11. Biểu đồ top 10 quốc gia có số khách hàng cao nhất và thấp nhất 50](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816172)

[Hình 3‑12. Biểu đồ doanh thu theo tháng và quý 50](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816173)

[Hình 3‑13. Biểu đồ doanh thu theo giờ 51](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816174)

[Hình 3‑14. Biểu đồ top 10 khách hàng dành nhiều tiền để mua nhất 52](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816175)

[Hình 3‑15. Biểu đồ tỷ lệ khách hàng mua 1 lần và nhiều lần 53](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816176)

[Hình 3‑16. Hiệu suất của 3 mô hình 56](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816177)

[Hình 3‑17. Biểu đồ giá trị mất mát của 3 mô hình trong huấn luyện 57](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816178)

[Hình 3‑18. Hiệu suất 3 mô hình với số lượng mẫu âm tính trên mỗi trường hợp dương tính 57](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816179)

[Hình 3‑19. Hiệu suất của NeuMF với các lớp ẩn khác nhau trong MLP 58](https://d.docs.live.net/5b79d32b0922df3f/Documents/IB/DEEP%20LEARNING%20VỚI%20HÀNH%20VI%20NGƯỜI%20TIÊU%20DÙNG/Nhóm%2011_Xây%20dựng%20hệ%20thống%20khuyến%20nghị%20cho%20người%20tiêu%20dùng%20dựa%20trên%20mô%20hình%20học%20sâu.docx#_Toc186816180)

# LỜI MỞ ĐẦU

## Lý do chọn đề tài

Trong thời đại công nghệ 4.0, bán hàng trực tuyến đã mang lại lợi nhuận lớn trong kinh doanh. Điều này là kết quả của sự phát triển mạnh mẽ của Internet. Bán hàng qua mạng đang trở nên phổ biến vì nó có khả năng cho khách hàng nhanh chóng tiếp cận hàng hóa và dịch vụ đa dạng.

Trên các sàn thương mại điện tử đã và đang thu hút sự quan tâm đáng kể của người tiêu dùng trên toàn thế giới. Theo thống kê của báo VietNamNet, doanh số trên sàn thương mại điện tử ở Việt Nam trong năm 2023 đạt gần 232,2 nghìn tỷ đồng, tăng 53,4% so với năm trước. Chỉ trong 9 tháng đầu năm 2024, doanh thu đạt hơn 227 nghìn tỷ đồng, với quý 3 đóng góp 84,75 nghìn tỷ đồng, tương đương 37% doanh thu hàng năm. Điều này cho thấy thị trường này có khả năng kiếm được rất nhiều tiền. Tuy nhiên sự đa dạng về sản phẩm cũng khiến người tiêu dùng khó lựa chọn.

Để người tiêu dùng có thể tiếp cận được với những sản phẩm đang phổ biến và phù hợp với mình, gợi ý đóng vai trò rất quan trọng. Một hệ thống gợi ý tốt có thể đóng vai trò như một người trung gian hỗ trợ khách hàng đưa ra các quyết định lựa chọn đúng đắn. Bằng cách xác định sở thích, nhu cầu kèm theo những thông tin cần thiết, hệ thống có thể đưa ra một tập các gợi ý giúp cho người tiêu dùng dễ dàng chọn lựa những sản phẩm. Qua đó hiệu suất của việc kinh doanh được tăng cao một cách đáng kể.

Mục tiêu chính của đề tài là tạo ra một hệ thống gợi ý sản phẩm cho người tiêu dùng dựa trên lịch sử mua hàng của họ để tối ưu hóa dự đoán khả năng ưu thích sản phẩm của người tiêu dùng và sau đó đề xuất các sản phẩm mà người tiêu dùng có thể quan tâm và ưu thích.

## Mục tiêu nghiên cứu

Tạo ra một hệ thống gợi ý sản phẩm giúp người tiêu dùng tìm được các sản phẩm phù hợp với sở thích cá nhân một cách nhanh chóng và hiệu quả, từ đó nâng cao trải nghiệm mua hàng của họ.

## Đối tượng và phạm vi nghiên cứu

* Đối tượng nghiên cứu:

+ Các mô hình học sâu ứng dụng trong hệ thống gợi ý

+ Dữ liệu hành vi người tiêu dùng

* Phạm vi nghiên cứu:

Nghiên cứu tập trung vào hệ thống gợi ý dựa trên lọc cộng tác, áp dụng kỹ thuật học sâu.

## Phương pháp nghiên cứu

* Phương pháp thu thập dữ liệu: Thu thập dữ liệu hành vi người tiêu dùng từ nguồn sẵn có
* Phương pháp phân tích: áp dụng các thuật toán học sâu để xây dựng hệ thống gợi ý, so sánh hiệu quả giữa các mô hình học sâu
* Công cụ hỗ trợ: Python, Pytorch, NumPy, Pandas, Matplotlib, Jupyter Notebook

## Kết cấu đề tài

Nội dung đề tài gồm các chương sau:

Chương 1: Tổng quan về kinh doanh thông minh

Chương 2: Học sâu và hệ thống gợi ý

Chương 3: Ứng dụng học sâu để xây dựng hệ thống gợi ý sản phẩm cho người tiêu dùng

# TỔNG QUAN KINH DOANH THÔNG MINH

## Khái niệm kinh doanh thông minh

### Định nghĩa

Kinh doanh thông minh (Business Intelligence – BI), là tập hợp phương pháp, công cụ và quy trình hiện đại mà các doanh nghiệp sử dụng để thu thập, phân tích thông tin và trực quan hóa những dữ liệu đó. Từ đó doanh nghiệp có thể đưa ra những quyết định kinh doanh hiệu quả, dựa trên cơ sở dữ liệu quá khứ và hiện tại cùng với dự báo về tương lai.

Mục tiêu chính của BI nhằm thúc đẩy doanh nghiệp đưa ra quyết định kinh doanh chính xác hơn, nhằm tăng trưởng doanh thu, tối ưu hóa hiệu quả hoạt động và tạo tiền đề để có thể cạnh tranh bền vững.

=> BI tận dụng công nghệ hiện đại để từ dữ liệu quá khứ và hiện tại.

### Đặc điểm chính

Từ khái niệm, ta có thể đặc điểm chính của BI như sau:

- Dựa trên dữ liệu quá khứ và hiện tại: để BI hoạt động một cách hiệu quả và đưa ra những phân tích có ý nghĩa, cần một lượng lớn data được thu thập từ nhiều thời điểm cũng như nhiều nguồn khác nhau.

- Phân tích dữ liệu: Sử dụng các kỹ thuật phức tạp và công cụ hiện đại, BI sẽ giúp doanh nghiệp phân tích, khám phá ra xu hướng, mối quan hệ tương quan,… của dữ liệu.

- Trực quan hóa: Từ những dữ liệu đã phân tích được, ta có thể đưa chúng lên biểu đồ, đồ thị để người xem dễ dàng nắm bắt được những thông tin quan trọng của dữ liệu một cách nhanh chóng và dễ dàng.

- Hỗ trợ ra quyết định: Những thông tin, số liệu và báo cáo BI cung cấp sẽ giúp lãnh đạo đưa ra những quyết định hiệu quả nhất.

- Dự báo: BI sử dụng các mô hình phân loại, dự báo để dự đoán xu hướng tương lai, giúp doanh nghiệp lên kế hoạch chuẩn bị, đón đầu xu thế.

### Một số công nghệ hỗ trợ BI

- Data Warehouse: đây là nơi lưu trữ dữ liệu từ nhiều nguồn, giúp doanh nghiệp dễ dàng truy cập và phân tích dữ liệu theo nhiều chiều khác nhau. Mục tiêu của data warehouse là giúp doanh nghiệp có góc nhìn toàn diện từ dữ liệu quá khứ và hiện tại, từ đó hỗ trợ việc phân tích, dự báo và đưa ra các quyết định chiến lược. Data thường được tổ chức dạng bảng, cho phép người dùng thực hiện các truy vấn phức tạp để khai thác dữ liệu.

- Enterprise resource Planning – ERP: là phần mềm quản lý toàn diện các hoạt động của doanh nghiệp, từ tài chính, nhân sự đến sản xuất và bán hàng. ERP sẽ cung cấp một lượng lớn dữ liệu chi tiết về hoạt động của doanh nghiệp.

- Query and report writing technologies: là những công cụ giúp doanh nghiệp truy xuất, phân tích và tạo báo cáo một cách trực quan từ data trong kho dữ liệu. Ta có thể kể tới 1 số công cụ thường gặp như: SQL, Tableau, Power BI.

- Data Mining and analytics tool: là những công cụ hỗ trợ việc tìm kiếm mẫu, xu hướng ẩn sâu trong dữ liệu, từ đó tìm ra những phát hiện mới, làm nền tảng cho việc dự đoán. Có thể kể tới như Helical Insight, Qlik Sense, Weka,…

- Decision support systems – DSS: là một phần mềm được sử dụng để hỗ trợ các quyết định, phán đoán và các phương án hành động cho doanh nghiệp. DSS sẽ sàng lọc, phân tích lượng dữ liệu khổng lồ, biên soạn thông tin toàn diện để có thể được sử dụng để giải quyết vấn đề trong quá trình ra quyết định.

- Customer relation management – CRM: là hệ thống quản lý cơ sở dữ liệu sẽ được lưu trữ, cập nhật thông tin khách hàng chính xác. CRM sẽ giúp doanh nghiệp phân tích, tạo danh sách khách hàng tiềm năng. Ngoài ra, doanh nghiệp còn xử lý được các vấn đề của khách hàng nhanh chóng và hiệu quả. Đồng thời đề ra những chiến lược chăm sóc khách hàng hợp lý, thiết lập mối quan hệ tốt với khách hàng. Từ đó tăng thiện cảm của họ với thương hiệu, tạo cơ hội phát triển công việc kinh doanh.

### Lợi ích

- BI có khả năng tổng hợp thông tin và giúp doanh nghiệp sử dụng nguồn thông tin đó hiệu quả hơn, dù cho có sự biến động của các yếu tố bên ngoài.

- Từ kết quả phân tích BI, doanh nghiệp nhìn ra hướng đi đúng đắn, đưa ra quyết sách nhanh chóng, kịp thời.

- Dự đoán xu hướng và hành vi khách hàng trong tương lai, mang đến cái nhìn tổng quan cho doanh nghiệp.

- Nhờ có BI mà doanh nghiệp có chiến lược phù hợp ứng với từng giai đoạn cụ thể.

- Giảm thiểu thời gian, tối ưu chi phí quản lý và vận hành doanh nghiệp.

- Ứng dụng tốt BI trong quản lý giúp gia tăng lợi thế khi cạnh tranh trên thị trường. Tạo tiền đề để tiếp cận nhiều cơ hội kinh doanh mới..

- Thông qua dữ liệu BI, doanh nghiệp biết được điểm yếu , điểm mạnh của mình. Từ đó, đưa ra giải pháp và thay đổi kịp thời, mang tính bền vững.

## Một số phương pháp và ứng dụng trong kinh doanh thông minh

### Phân tích dữ liệu lớn

Phân tích dữ liệu lớn đóng vai trò cốt lõi trong việc nâng cao khả năng xử lý dữ liệu khổng lồ và đa dạng của BI. Công nghệ này cho phép doanh nghiệp thu thập, lưu trữ, và phân tích các tập dữ liệu từ nhiều nguồn khác nhau, có thể kể tới như: mạng xã hội, giao dịch trực tuyến, và cảm biến IoT. Cùng với những công cụ mạnh mẽ như Hadoop và Apache Spark, big data giúp BI không chỉ dừng lại ở việc báo cáo mà còn khám phá các xu hướng ẩn, dự đoán hành vi khách hàng, và phát hiện gian lận. Sức mạnh của dữ liệu lớn giúp doanh nghiệp có cái nhìn toàn diện, hỗ trợ ra quyết định chính xác và kịp thời.

### Trí tuệ nhân tạo

Trí tuệ nhân tạo (AI) đưa BI lên một tầm cao mới, từ việc phân tích dữ liệu thụ động, hoàn toàn dựa vào con người sang việc tự động hóa và đưa ra dự đoán thông minh. AI có thể sử dụng các thuật toán học máy (Machine Learning) để dự đoán xu hướng thị trường, phân khúc khách hàng, và tối ưu hóa chiến lược kinh doanh. Đồng thời, với công nghệ xử lý ngôn ngữ tự nhiên (NLP), AI cho phép BI phân tích phản hồi khách hàng từ văn bản, email hoặc mạng xã hội, từ đó giúp doanh nghiệp hiểu sâu hơn về cảm xúc và nhu cầu của người dùng. Nhờ tích hợp AI, BI trở thành một công cụ chiến lược, không chỉ phân tích dữ liệu mà còn hỗ trợ đưa ra quyết định một cách linh hoạt và sáng tạo.

### Internet vạn vật

Internet vạn vật (IoT) đã mở ra một nguồn dữ liệu thời gian thực đầy tiềm năng cho BI, cung cấp thông tin từ các thiết bị kết nối có thể kể tới như: cảm biến, camera, và thiết bị thông minh. Dữ liệu từ IoT giúp doanh nghiệp theo dõi tình hình hoạt động của hệ thống, phân tích hành vi khách hàng trong thời gian thực, hoặc dự đoán các vấn đề bảo trì cho thiết bị. Ví dụ, trong ngành bán lẻ, dữ liệu từ IoT có thể giúp phân tích luồng khách hàng trong cửa hàng để tối ưu hóa bố trí sản phẩm. Khi tích hợp IoT, BI không chỉ làm việc với dữ liệu lịch sử mà còn mang lại khả năng giám sát và phân tích ngữ cảnh thời gian thực, từ đó giúp nâng cao hiệu suất vận hành.

### Điện toán đám mây

Điện toán đám mây là một nền tảng hỗ trợ mạnh mẽ cho BI, cung cấp khả năng lưu trữ linh hoạt và xử lý dữ liệu nhanh chóng. BI trên đám mây cho phép doanh nghiệp truy cập dữ liệu mọi lúc, mọi nơi, đồng thời giảm chi phí đầu tư và bảo trì hệ thống thay vì phụ thuộc vào cơ sở hạ tầng tại chỗ như trước đây. Các nền tảng phổ biến ngày nay có thể kể tới như: AWS, Google Cloud, và Microsoft Azure không chỉ cung cấp không gian lưu trữ mà còn tích hợp các công cụ phân tích dữ liệu lớn và trí tuệ nhân tạo, giúp BI trở thành một giải pháp toàn diện. Bên cạnh đó, điện toán đám mây còn đảm bảo khả năng mở rộng, phù hợp với nhu cầu ngày càng tăng của doanh nghiệp, từ đó thúc đẩy quá trình ra quyết định hiệu quả hơn.

### Ứng dụng

BI sử dụng dữ liệu từ quá khứ và hiện tại để đưa ra dự báo về tương lai. Từ đó, ta có thể đến một số ứng dụng tiêu biểu của kinh doanh thông minh như sau:

- Dự báo nhu cầu khách hàng:

Kinh doanh thông minh hỗ trợ doanh nghiệp dự báo chính xác nhu cầu của khách hàng thông qua việc phân tích dữ liệu lịch sử, xu hướng thị trường và hành vi tiêu dùng. Bằng các thuật toán tiên tiến, BI giúp doanh nghiệp hiểu rõ thời điểm và những gì khách hàng cần, từ đó tối ưu hóa chiến lược cung ứng, giảm chi phí vận chuyển, tình trạng hàng tồn kho hoặc thiếu hụt. Điều này đặc biệt hữu ích trong ngành bán lẻ, sản xuất và các lĩnh vực có chu kỳ biến động nhu cầu cao.

- Tối ưu hóa hoạt động kinh doanh:

Từ việc cung cấp thông tin chi tiết về hiệu suất hoạt động trong doanh nghiệp, BI giúp doanh nghiệp xác định các quy trình thiếu hiệu quả, gây lãng phí tài nguyên. Dựa trên các dữ liệu được phân tích, doanh nghiệp có thể tìm ra những vị trí, vai trò hoạt động chưa hiệu quả, tinh chỉnh các hoạt động như sản xuất, phân phối, hoặc quản lý nhân sự, từ đó giảm chi phí và tăng năng suất. Ví dụ, trong ngành logistics, BI có thể giúp tối ưu hóa tuyến đường vận chuyển để tiết kiệm thời gian và chi phí.

- Quản lý rủi ro:

Một ứng dụng quan trọng khác của BI chính là khả năng phát hiện và giảm thiểu rủi ro trong kinh doanh. BI giúp doanh nghiệp phân tích dữ liệu lịch sử để dự đoán các rủi ro tiềm ẩn như gian lận tài chính, sự cố chuỗi cung ứng, hoặc thay đổi thị trường. Nhờ vậy, các nhà quản lý có thể đưa ra các biện pháp phòng ngừa kịp thời, giảm thiểu thiệt hại và bảo vệ uy tín doanh nghiệp.

- Phát triển sản phẩm và dịch vụ mới:

Việc doanh nghiệp hiểu rõ nhu cầu của thị trường và khách hàng cũng thúc đẩy việc phát triển sản phẩm mới để đáp ứng như cầu đó. Với những thông tin đó, doanh nghiệp có thể đưa ra các thiết kế sản phẩm phù hợp thị hiếu, đáp ứng và tăng khả năng cạnh tranh trên thị trường.

- Tăng cường trải nghiệm khách hàng:

Thông qua hành vi mua sắm, lịch sử giao dịch hoặc mạng xã hội, doanh nghiệp có thể hiểu rõ hơn về sở thích và thói quen tiêu dùng của khách hàng. Điều đó cung cấp cho doanh nghiệp khả năng cải thiện dịch vụ như hậu mãi và xây dựng mối quan hệ chặt chẽ hơn với khách hàng.

=> BI là một công cụ không thể thiếu đối với doanh nghiệp, hỗ trợ doanh nghiệp duy trì lợi thế cạnh tranh và phát triển bền vững.

## Những thách thức trong kinh doanh thông minh

- Khó khăn trong việc thu thập và sử lý dữ liệu:

Trong công cuộc chuyển đổi số hiện nay, lượng dữ liệu mà doanh nghiệp cần xử lý ngày càng lớn, phức tạp và đa dạng, chúng có thể đến từ nhiều nguồn như giao dịch, mạng xã hội, hoặc thiết bị IoT. Tuy nhiên, việc thu thập và làm sạch dữ liệu thường gặp khó khăn do dữ liệu không đồng nhất, thiếu chính xác hoặc không đầy đủ. Điều này có thể dẫn đến kết quả phân tích sai lệch, ảnh hưởng tiêu cực đến các quyết định kinh doanh.

- Rào cản công nghệ:

BI cần rất nhiều công nghệ hiện đại để phát huy tối đa khả năng. Tuy nhiên, việc đầu tư và duy trì các hệ thống này đòi hỏi nguồn lực tài chính lớn, đặc biệt là đối với các doanh nghiệp vừa và nhỏ khó lòng có thể kham nổi. Ngoài ra, thế giới phát triển không ngừng, tốc độ thay đổi nhanh của công nghệ cũng đặt ra thách thức trong việc theo kịp xu hướng và nâng cấp hệ thống thường xuyên.

- Vấn đề an toàn bảo mật dữ liệu:

Vấn đề này liên quan đến gốc rễ của BI – dữ liệu. Có thể kể tới một số rủi ro như: tấn công mạng, rò rỉ thông tin,… ít nhiều sẽ gây tổn hại đến danh tiếng, thậm chí là khả năng hoạt động của doanh nghiệp.

- Thiếu hụt nhân lực:

Việc sử dụng hàng loạt công nghệ và quy trình hiện đại đòi hỏi nhân lực có trình độ kỹ năng chuyên môn cao trong lĩnh vực phân tích cũng như học máy. Tuy nhiên ở Việt Nam, một thị trường mới nổi đây là bài toán mà cần thời gian mới có thể đáp ứng đủ được.

## Các bước cơ bản trong kinh doanh thông minh

- Xác định mục tiêu kinh doanh:

Doanh nghiệp cần trả lời những câu hỏi quan trọng như: “Chúng ta muốn giải quyết vấn đề gì?", "Cần đạt được điều gì qua phân tích dữ liệu?". Việc đặt ra mục tiêu cụ thể giúp định hướng các bước tiếp theo và đảm bảo rằng mọi nỗ lực đều tập trung vào việc mang lại giá trị thực tiễn.

- Thu thập và xử lý dữ liệu:

Thông qua các nguồn như CRM, mạng xã hội, cảm biến IoT và các giao dịch tài chính, doanh nghiệp sẽ thu được lượng lớn data thô. Dữ liệu này cần được làm sạch, chuẩn hóa để đảm bảo tính nhất quán và chính xác. Đây là quá trình tối quan trọng vì chất lượng dữ liệu quyết định hiệu quả và độ chính xác của các mô hình phân tích và dự đoán.

- Phân tích dữ liệu:

Sau khi dữ liệu được xử lý, ta sẽ thu được nhiều thông tin quan trọng của dữ liệu, từ đó kế hợp bảng biểu trực quan hóa cùng kiến thức chuyên ngành, doanh nghiệp sẽ nhận ra được xu hướng, mẫu hình hoặc các vấn đề cần chú ý, từ đó đưa ra các thông tin hữu ích cho giai đoạn ra quyết định chiến lược.

- Xây dựng mô hình học sâu:

Doanh nghiệp có thể xây dựng mô hình học sâu để phân tích dữ liệu ở mức độ cao hơn để giải quyết những mục tiêu có độ phức tạp cao như dự đoán như cầu khách hàng, phát hiện gian lận tài chính,… những mô hình này cần có khả năng học hỏi từ big data và cải thiện kết quả dự đoán liên tục.

- Đánh giá và triển khai mô hình:

Sau khi xây dựng xong mô hình, ta cần kiểm tra và đánh giá để đảm bảo mô hình có thể đáp ứng yêu cầu của doanh nghiệp, nếu mô hình có tính chính xác cao và đáng tin cậy thì sẽ được triển khai vận hành như một công cụ hỗ trợ ra quyết định.

- Theo dõi và điều chỉnh :

Sau khi triển khai mô hình, doanh nghiệp vẫn cần liên tục theo dõi hiệu quả hoạt động trong thực tế của mô hình, trong trường hợp có sự xuất hiện của biến số mới, mô hình cần được điều chỉnh để duy trì độ chính xác và giá trị ứng dụng.

# HỌC SÂU VÀ HỆ THỐNG GỢI Ý

## Mạng nơ-ron (Neural network)

### Giới thiệu về học sâu

#### Khái niệm

Lĩnh vực trí tuệ nhân tạo (Artificial Intelligence – AI) và học máy (Machine Learning – ML) đang phát triển với tốc độ chóng mặt và bùng nổ, tạo ra cả nỗi lo ngại và sự hứng khởi toàn cầu. Nếu như nhiều người có cái nhìn tổng quan về học máy và trí tuệ nhân tạo, thì học sâu (Deep Learning – DL) – một nhánh chuyên biệt của học máy lại là một khái niệm phức tạp, đòi hỏi mức độ giải thích sâu sắc và kỹ lưỡng hơn.

Học sâu là tập con của học máy, tập trung vào sử dụng mạng nơ-ron nhân tạo (Artificial Neural Networks – ANNs) để thực hiện các nhiệm vụ như phân loại, hồi quy, và học biểu diễn đặc trưng. Học sâu tập trung vào việc sắp xếp các nơ-ron nhân tạo thành nhiều lớp, sau đó “huấn luyện” chúng để xử lý dữ liệu một cách hiệu quả. Tính từ “deep” trong Deep Learning không nói về việc hiểu biết sâu sắc hơn mà ám chỉ việc sử dụng nhiều lớp biểu diễn trong mạng nơ-ron, có thể dao động từ ba đến hàng trăm, thậm chí hàng nghìn lớp. Các phương pháp được áp dụng trong học sâu có thể là học có giám sát, bán giám sát hoặc không giám sát.

Các thuật toán học sâu được xây dựng với các lớp được kết nối:

- Lớp đầu vào (Input Layer): Gồm một số nút để nhập dữ liệu đầu vào

- Lớp ẩn (Hidden Layers): Mỗi lớp ẩn bao gồm các nơ-ron thần kinh đóng vai trò trung gian, xử lý và truyền tải thông tin giữa đầu vào và đầu ra của mạng. Mỗi nơ-ron trong lớp ẩn nhận tín hiệu từ các nơ-ron của lớp trước, xử lý thông qua trọng số, độ lệch và hàm kích hoạt, sau đó truyền tín hiệu đến lớp kế tiếp. Cường độ tín hiệu mà nơ-ron gửi đi được quyết định bởi sự tương tác giữa các yếu tố này, giúp mạng học được các mẫu phức tạp và trừu tượng từ dữ liệu.

- Lớp đầu ra (Output Layer): Gồm các nút xuất dữ liệu của mô hình

Mô hình học sâu có thể nhận diện nhiều hình mẫu phức tạp trong hình ảnh, văn bản, âm thanh và các loại dữ liệu khác để tạo ra thông tin phân tích chuyên sâu và dự đoán chính xác. Điều đó giúp chúng ta ứng dụng các phương pháp học sâu để tự động hóa các tác vụ thường đòi hỏi trí tuệ con người, chẳng hạn như mô tả hình ảnh hoặc chép lời một tập tin âm thanh. Ngoài ra, các mô hình học sâu còn được ứng dụng rộng rãi trong nhiều lĩnh vực khác như y tế, học sâu đã mang lại những đột phá quan trọng trong hỗ trợ chẩn đoán bệnh thông qua dữ liệu y khoa từ đó đưa ra những đề xuất kế hoạch điều trị phù hợp. Trong tài chính, học sâu có khả năng dự đoán xu hướng thị trường tài chính giúp các nhà đầu tư và chuyên gia tài chính đưa ra những quyết định sáng suốt đồng thời phát hiện các giao dịch gian lận. Trong lĩnh vực thương mại điện tử, học sâu giúp các doanh nghiệp cải thiện trải nghiệm khách hàng bằng cách đưa ra các gợi ý sản phẩm phù hợp dựa trên hành vi của người tiêu dùng. … Với tiềm năng tự động hóa và học hỏi từ dữ liệu lớn, học sâu đang trở thành một công cụ không thể thiếu trong việc giải quyết các bài toán phức tạp, mang lại giá trị đột phá trong nhiều ngành công nghiệp trên toàn thế giới.

#### Ưu nhược điểm

* Ưu điểm:

Học sâu là một bước ngoặt lớn trong lĩnh vực trí tuệ nhân tạo, cho phép các nhà khoa học dữ liệu xây dựng nhiều mô hình có độ chính xác cao trong các lĩnh vực nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, …. Một số ưu điểm nổi bật của học sâu bao gồm:

+ Tự động học và cải tiến tính năng:

Các thuật toán học sâu có khả năng tự động học các tính năng từ dữ liệu mà không cần sự can thiệp thủ công.

+ Xử lý dữ liệu lớn và phức tạp:

Các thuật toán học sâu vượt trội trong việc phân tích lượng lớn dữ liệu phức tạp từ nhiều nguồn và đưa ra những kết luận sâu sắc sau quá trình.

+ Nâng cao hiệu suất:

Thuật toán học sâu đã được chứng minh là có hiệu suất xuất sắc so với các phương pháp truyền thống trên nhiều vấn đề, bao gồm nhận diện hình ảnh và giọng nói, xử lý ngôn ngữ tự nhiên và thị giác máy tính.

+ Xử lý các mối quan hệ phi tuyến tính:

Học sâu tìm ra các mối quan hệ phi tuyến tính trong dữ liệu đầu vào mà các phương pháp truyền thống khó thực hiện được, giúp các mẫu dữ liệu phức tạp được thể hiện chính xác hơn, đồng thời xử lý và mở rộng tính ứng dụng của học sâu trong nhiều lĩnh vực khác.

+ Xử lý dữ liệu tuần tự:

Các thuật toán deep learning chuyên sâu như Mạng nơron hồi quy (Recurrent Neural Networks – RNN) và Bộ nhớ dài-ngắn hạn (Long Short-term Memory – LSTM) rất thành thạo trong việc xử lý dữ liệu tuần tự như chuỗi thời gian, lời nói và văn bản. Khả năng duy trì ngữ cảnh và bộ nhớ theo thời gian cho phép chúng đưa ra các dự đoán thông tin dựa trên dữ liệu đầu vào lịch sử.

* Nhược điểm:

Bên cạnh những lợi ích to lớn, học sâu vẫn tồn tại một số hạn chế cần phải xem xét cẩn thận:

+ Chi phí điện toán cao:

Huấn luyện các mô hình học sâu đòi hỏi lượng lớn tài nguyên tính toán, bao gồm GPU mạnh và thời gian huấn luyện dài.

+ Sự quá khớp (Overfiting):

Hiện tượng này rất phổ biến trong học sâu, đặc biệt là đối với mạng nơ-ron lớn và xuất phát từ nhiều yếu tố như dữ liệu bị thiếu, mô hình phức tạp hoặc thiếu chuẩn hóa.

+ Thiếu khả năng diễn giải:

Đa phần các mô hình học sâu có nhiều layer có thể phức tạp và khó diễn giải làm cho quá trình dự đoán và xác định lỗi hoặc sai lệch trở nên khó khăn.

+ Giới hạn đối với dữ liệu được đào tạo:

Các mô hình học sâu chỉ đưa ra được dự đoán dựa trên dữ liệu được đào tạo, hạn chế khả năng bao quát cho các tình huống hoặc bối cảnh mới mà chưa được trình bày trong dữ liệu đào tạo của mô hình.

+ Mô hình hộp đen:

Một số mô hình học sâu được coi là “hộp đen” vì rất khó để hiểu được cách mô hình đưa ra dự đoán và xác định các yếu tố ảnh hưởng đến dự đoán đó.

#### Các loại mạng học sâu phổ biến

Các mô hình học sâu có thể tự động học các tính năng từ nguồn dữ liệu lớn dựa vào quá trình sử dụng mạng nơ-ron nhân tạo (ANN) với nhiều lớp ẩn. Các loại mạng nơ-ron được sử dụng rộng rãi trong học sâu gồm:

* Mạng nơ-ron truyền thẳng (Feed-forward Neural Network – FNN):

FNN là loại mạng nơ-ron ra đời sớm nhất và đơn giản nhất. Trong mạng này, thông tin chỉ di duyển một chiều từ các node đầu vào, thông qua các node ẩn và đi đến các node đầu ra mà không trải qua chu trình hay vòng lặp vào trong mạng. FNN thuận tiện cho việc lập trình lại và xử lý các câu hỏi phi tuyết tính, đã được sử dụng rộng rãi cho các tác vụ như phân loại hình ảnh, nhận dạng giọng nói và xử lý ngôn ngữ tự nhiên.

* Mạng nơ-ron tích chập (Convolutional Neural Network – CNN):

CNN là một trong những mô hình để nhận dạng là phân loại hình ảnh, đặc biệt được ứng dụng rộng rãi trong việc xác định đối tượng và nhận dạng khuôn mặt. CNN có khả năng tự động học các tính năng từ hình ảnh, giúp chúng phù hợp với các tác vụ như phân loại hình ảnh, phát hiện đối tượng và phân đoạn hình ảnh.

* Mạng nơ-ron hồi quy (Recurrent Neural Network – RNN):

RNN – là một loại mạng nơ-ron có khả năng xử lý dữ liệu tuần tự, chẳng hạn như chuỗi thời gian và ngôn ngữ tự nhiên. RNN có khả năng duy trì trạng thái nội bộ để lưu giữ thông tin về các đầu vào trước đó, khiến chúng phù hợp với các tác vụ như nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên và dịch ngôn ngữ.

### Mạng nơ-ron truyền thẳng

Mạng thần kinh truyền thẳng hay mạng nơ-ron truyền thẳng (Feedforward Neural Network – FNN) là một mạng thần kinh nhân tạo trong đó kết nối giữa các nút không hình thành một chu trình. Trong mạng này, thông tin di chuyển chỉ một chiều hướng đến từ các nút đầu vào, thông qua các nút ẩn (nếu có) và đi đến các nút đầu ra. Không có chu trình (chu kỳ) hoặc vòng lặp trong mạng.

FNN có ưu điểm là đơn giản, dễ hiểu và dễ huấn luyện. Tuy nhiên, FNN không có cơ chế ghi nhớ trạng thái trước đó như mạng nơ-ron hồi quy (RNN) nên chủ yếu được sử dụng cho học có giám sát trong các trường hợp dữ liệu cần học không có tính tuần tự hoặc phụ thuộc vào thời gian (như chuỗi thời gian, văn bản). Dựa trên số lượng lớp ẩn và cách kết nối giữa các nơ-ron, có thể chia FNN thành hai loại:

- Mạng đơn lớp: Chỉ có một lớp neuron, thường dùng cho các bài toán đơn giản như phân loại tuyến tính.

- Mạng đa lớp: Có nhiều hơn một lớp ẩn, cho phép mạng học được các hàm phi tuyến phức tạp hơn.

Để đơn giản và tránh hiểu nhầm, kiến trúc mạng truyền thẳng được nói đến trong bài này là mạng nơ-ron tuyến tính nhiều lớp ( Multi-layer Perceptron – MLP)

#### Giới thiệu về mạng nơ-ron tuyến tính nhiều lớp (MLP)

Trong học sâu, mạng perceptron nhiều lớp (Multi-layer Perceptron - MLP) là một thuật ngữ chỉ các mạng nơ-ron truyền thẳng hiện đại. Mạng này bao gồm các nơ-ron được kết nối hoàn toàn với nhau, sử dụng các hàm kích hoạt phi tuyến, được tổ chức thành từng lớp. Điểm nổi bật của MLP là được ứng dụng rộng rãi trong phân biệt dữ liệu không thể phân tách tuyến tính.

Các mạng nơ-ron hiện đại được huấn luyện bằng thuật toán lan truyền ngược (backpropagation). MLP ra đời từ nỗ lực cải tiến các perceptron một lớp – vốn chỉ xử lý được dữ liệu có thể phân tách tuyến tính. Trước đây, perceptron truyền thống sử dụng hàm bước Heaviside làm hàm kích hoạt phi tuyến. Tuy nhiên, để thuật toán lan truyền ngược hoạt động hiệu quả, các MLP hiện đại sử dụng các hàm kích hoạt liên tục như sigmoid hoặc ReLU. MLP là nền tảng của học sâu và có thể được ứng dụng trong nhiều lĩnh vực đa dạng.

MLP là một loại mạng nơ-ron nhân tạo được ứng dụng rộng rãi trong nhiều lĩnh vực nhờ khả năng mô hình hóa các mối quan hệ phi tuyến tính: xử lý dữ liệu phi cấu trúc (hình ảnh, văn bản, âm thanh); phân tích kinh doanh với dự báo tài chính, phân tích hành vi khách hàng và phát hiện gian lận; phân tích hình ảnh y tế và dự đoán tiến triển bệnh. MLP còn được áp dụng trong các lĩnh vực như kỹ thuật (tối ưu hóa quy trình sản xuất, mô phỏng vật lý, dự báo thời tiết), giao thông (điều hướng xe tự hành, dự báo lưu lượng giao thông), game và giải trí (phát triển AI trong game, tạo nội dung giải trí). Ngoài ra MLP có thể được kết hợp để xây dựng hệ thống gợi ý (Recommendation System) nhờ vào khả năng mô hình hóa các mối quan hệ phi tuyến tính giữa người dùng và các sản phẩm. Trong hệ thống gợi ý, MLP thường được sử dụng để học các đặc trưng tiềm ẩn giữa người dùng và các sản phẩm hoặc dịch vụ mà họ tương tác, giúp đưa ra những gợi ý chính xác và phù hợp; đồng thời kết hợp với các kỹ thuật khác như Collaborative Filtering, Content-Based Filtering, hoặc Matrix Factorization để cải thiện hiệu quả hệ thống gợi ý, giúp mang lại trải nghiệm người dùng tốt hơn và tăng tỷ lệ chuyển đổi trong thương mại điện tử hoặc các nền tảng trực tuyến.

#### Kiến trúc của MLP

Trong mạng MLP, ngoài lớp Input (lớp đầu vào) và Output (lớp đầu ra), các lớp nơ-ron ở giữa được gọi chung là lớp Hidden (lớp ẩn). Chính nhờ tính chất phi tuyến tính này, MLP có khả năng học các hàm phi tuyến tính phức tạp và phân biệt dữ liệu không thể phân tách tuyến tính.

* Lớp đầu vào:

+ Lớp đầu vào đóng vai trò là điểm nhập dữ liệu vào mạng nơ-ron.

+ Mỗi nơ-ron trong lớp đầu vào tương ứng với một tính năng hoặc biến đầu vào trong tập dữ liệu, và được đánh trọng số (weights) riêng biệt.

+ Độ đa chiều của dữ liệu đầu vào quyết định số lượng nơ-ron trong lớp đầu vào.

* Lớp ẩn:

+ Các lớp ẩn là lớp trung gian giữa lớp đầu vào và lớp đầu ra, nơi diễn ra các phép tính và chuyển đổi phức tạp.

+ Mỗi nơ-ron trong lớp ẩn nhận dữ liệu đầu vào từ các nơ-ron ở lớp trước đó và tạo ra đầu ra dựa trên tổng trọng số của các dữ liệu đầu vào này.

+ Số lượng lớp ẩn và số lượng nơ-ron trong mỗi lớp ẩn là các siêu tham số có thể cấu hình được, có thể thay đổi tùy theo độ phức tạp của vấn đề và đặc điểm của dữ liệu.

* Lớp đầu ra:

+ Lớp đầu ra có trách nhiệm tạo ra đầu ra cuối cùng của mạng nơ-ron.

+ Số lượng nơ-ron trong lớp đầu ra phụ thuộc vào bản chất của nhiệm vụ dự đoán.

+ Đối với các tác vụ phân loại nhị phân, một nơ-ron đơn lẻ có hàm kích hoạt hình sigmoid có thể được sử dụng để tạo ra điểm xác suất.

+ Đối với các tác vụ phân loại đa lớp , lớp đầu ra có thể bao gồm nhiều nơ-ron với hàm kích hoạt softmax để đưa ra xác suất lớp đầu ra.

+ Trong các tác vụ hồi quy, lớp đầu ra thường chứa một nơ-ron duy nhất có hàm kích hoạt tuyến tính để tạo ra các dự đoán có giá trị liên tục.

Mỗi nơ-ron trong MLP được kết nối với mọi nơ-ron ở lớp tiếp theo thông qua các kết nối có trọng số. Các trọng số này xác định sức mạnh của các kết nối giữa các nơ-ron và chi phối ảnh hưởng của từng đầu vào đến đầu ra của nơ-ron. Trong quá trình đào tạo, trọng số được điều chỉnh thông qua quá trình truyền ngược để giảm thiểu lỗi dự đoán và cải thiện hiệu suất của mạng.

Trong MLP, dữ liệu chảy từ lớp đầu vào đến lớp đầu ra, đi qua các lớp ẩn, và mỗi nơ-ron tính toán đầu ra dựa trọng số, đầu vào, và hàm kích hoạt qua các lớp. Công thức tổng quát cho đầu ra của một lớp (layer):

*z(l) =* *W(l) \* a(l − 1) + b(l)*

*a(l) = f(l)( z(l))*

Trong đó:

*l*: số thứ tự lớp cần tính đầu ra (*l = 1, 2, …, L*)

*z(l)*: tổng trọng số của các đầu vào đến lớp *l*

*W(l)*: ma trận weight (ma trận trọng số) của lớp *l* (kích thước = số node lớp *l* số node lớp *l – 1* )

*b(l)*: vector bias của lớp *l* (kích thước = số node lớp *l*)

*a(l-1)*: đầu ra của lớp *l – 1*

*a(l)*: đầu ra của lớp *l*

*f(l)*: hàm kích hoạt (activation function) tại lớp *l* (tùy vào đầu ra của bài)

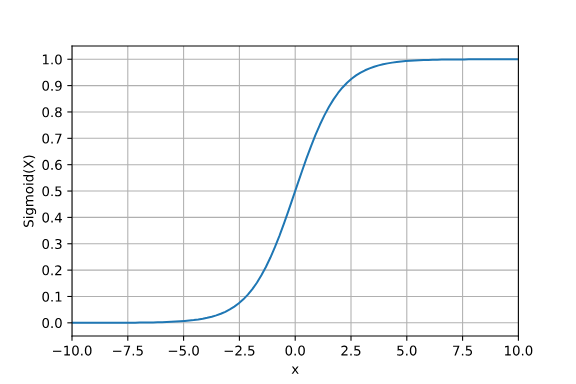
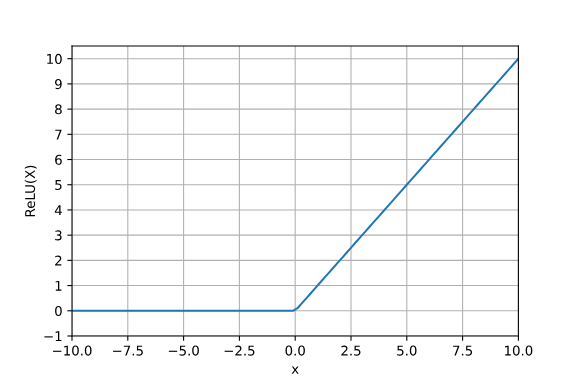
*+*: phép cộng vector/matrix tương ứng.

Chú ý:

- Tầng đầu tiên: *a0 = x* (tức là đầu vào ban đầu)

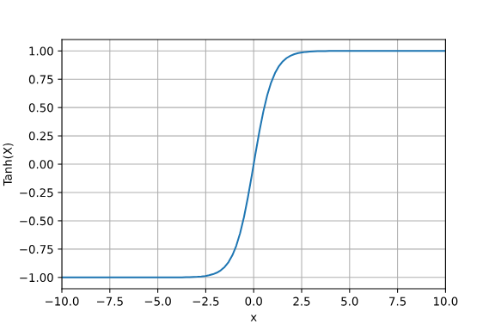
- Hàm kích hoạt: Tùy thuộc vào bài toán, mỗi tầng có thể sử dụng các hàm kích hoạt khác nhau:

+ ReLU: *f(x) = max(0,z)*, ReLU đơn giản lọc các giá trị < 0.

**+ Sigmoid*: f(x) =*  , nhận đầu vào là một số thực và chuyển thành một giá trị trong khoảng (0;1).

Hình 2‑1. Hàm Sigmoid

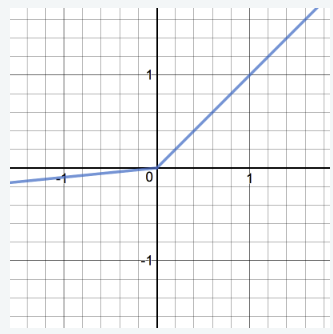
Hình 2‑2. Hàm ReLU

+ Tanh: *f(x) =*  , nhận đầu vào là một số thực và chuyển thành một giá trị trong khoảng (-1; 1).

Hình 2‑3. Hàm Tanh

+ Leaky ReLU: *f(x) =*  , là một cố gắng trong việc loại bỏ "dying ReLU". Thay vì trả về giá trị 0 với các đầu vào < 0 thì Leaky ReLU tạo ra một đường xiên có độ dốc nhỏ.

Hình 2‑4. Hàm Leaky ReLU

+ Softmax (cho bài toán phân loại): f(x) =

### Thuật toán lan truyền ngược (Backpropagation)

Backpropagation (viết tắt của “backward propagation of errors”) là một thuật toán cơ bản dùng để huấn luyện mạng nơ-ron bằng cách điều chỉnh các tham số của mạng để tối ưu hóa hàm mất mát (Lost function) hay giảm thiểu lỗi dự đoán.

Backpropagation là một thuật toán mạnh mẽ trong học sâu, chủ yếu được sử dụng để đào tạo mạng nơ-ron nhân tạo, đặc biệt là mạng truyền thẳng.

#### Quy trình thực hiện thuật toán lan truyền ngược

**Giai đoạn 1: Lan truyền**

*Bước 1: Lan truyền thẳng (Forward Propagation)*

Đầu vào được lan truyền thông qua từng lớp của mạng nơ-ron, qua các hàm kích hoạt để tạo đầu ra.

* Công thức tính đầu ra của một lớp:

*z(l) =* *W(l) \* a(l − 1) + b(l)*

*a(l) = f(l)( z(l))*

Trong đó:

*l*: số thứ tự lớp cần tính đầu ra (*l = 1, 2, …, L*)

*z(l)*: tổng trọng số của các đầu vào đến lớp *l*

*W(l)*: ma trận weight (ma trận trọng số) của lớp *l*

*b(l)*: vector bias của lớp *l*

*a(l-1)*: đầu ra của lớp *l – 1*

*a(l)*: đầu ra của lớp *l*

*f(l)*: hàm kích hoạt (activation function) tại lớp *l*

* Tại đầu ra của mạng , chúng ta có đầu ra dự đoán

= (*L* – số lớp tổng cộng)

*Bước 2: Lan truyền ngược (Backforward Propagation)*

Truyền ngược của các kích hoạt đầu ra thông qua mạng lưới nơ-ron sử dụng mục tiêu huấn luyện mô hình để tạo ra các delta dùng để cập nhật.

* Tính toán mất mát (so sánh đầu ra dự đoán của mạng với đầu ra mục tiêu)

Công thức tính hàm mất mát (Loss fuction):

Trong đó:

– Hàm mất mát trên toàn bộ dữ liệu

– số mẫu trong tập huấn luyện

– Hàm mất mát tại mỗi điểm (, )

– Dữ liệu đầu vào

– Giá trị đầu ra mong muốn

– Giá trị đầu ra dự đoán

* Tính delta tại lớp cuối:

Trong đó:

– Vector sai số của lớp cuối cùng

– Gradient của hàm mất mát theo đầu ra của lớp cuối cùng

– Đạo hàm của hàm kích hoạt tại lớp cuối cùng

* Lan truyền ngược để tính delta cho các lớp trước đó:

Trong đó:

– Vector sai số của lớp *l*

– Vector sai số của lớp *l + 1*

– Ma trận trọng số của lớp *l + 1* chuyển vị

– Đạo hàm của hàm kích hoạt tại lớp *l*

* Gradient của trọng số và bias:

+ Gradient của trọng số:

+ Gradient của bias:

**Giai đoạn 2: Cập nhật trọng số**

Nhân các delta đầu ra và hàm kích hoạt đầu vào để có được gradient của trọng số của nó. Trừ một tỷ lệ (%) từ gradient của trọng số (hệ số học alpha của trọng số).

Lặp lại giai đoạn 1 và 2 cho đến khi hàm mất mát đạt giá trị nhỏ mất hoặc số vòng lặp đạt tới giới hạn.

* Công thức cập nhật:

+ Cập nhật trọng số:

+ Cập nhật bias:

Trong đó:

– Hệ số học (Learning rate)

#### Vai trò của thuật toán lan truyền ngược

Thuật toán lan truyền ngược đóng vai trò rất quan trọng trong việc giúp các mạng nơ-ron cải thiện theo thời gian:

* Cập nhật trọng số hiệu quả:

Lan truyền ngược tính toán đạo hàm của hàm mất mát đối với từng trọng số bằng cách sử dụng quy tắc chuỗi, giúp cập nhật trọng số một cách hiệu quả.

* Khả năng mở rộng:

Thuật toán lan truyền ngược có thể mở rộng tốt với các mạng có nhiều lớp và kiến trúc phức tạp, làm cho học sâu trở nên khả thi.

* Học tự động:

Với lan truyền ngược, quá trình học trở nên tự động, và mô hình có thể điều chỉnh bản thân để tối ưu hóa hiệu suất.

#### Các biến thể của thuật toán lan truyền ngược

Backpropagation có một số biến thể và cải tiến nhằm cải thiện hiệu suất và khắc phục một số hạn chế, bao gồm:

- Stochastic Gradient Descent (SGD):

Trong khi phương pháp truyền thống của Gradient Descent cập nhật trọng số sau mỗi lần tính toán toàn bộ bộ dữ liệu (batch), SGD cập nhật trọng số sau mỗi mẫu (sample) hoặc sau mỗi batch con. Điều này giúp giảm thiểu thời gian tính toán và có thể giúp mạng hội tụ nhanh hơn.

- Mini-batch Gradient Descent:

Là một biến thể của SGD, trong đó bộ dữ liệu được chia thành các mini-batch nhỏ hơn và mô hình sẽ cập nhật trọng số sau mỗi mini-batch. Phương pháp này cân bằng giữa tốc độ và độ chính xác của quá trình học.

- Momentum:

Momentum giúp vượt qua các điểm cục bộ trong quá trình tối ưu hóa bằng cách "đẩy" trọng số theo hướng của gradient. Điều này làm giảm sự dao động và giúp quá trình huấn luyện hội tụ nhanh hơn.

- Adam (Adaptive Moment Estimation):

Là một biến thể nâng cao của Gradient Descent, Adam sử dụng thông tin về gradient đầu tiên và gradient thứ hai để điều chỉnh tốc độ học của mỗi trọng số một cách tự động, giúp quá trình học ổn định hơn và nhanh chóng hội tụ.

- RMSProp (Root Mean Square Propagation):

RMSProp điều chỉnh tốc độ học cho mỗi trọng số dựa trên bình phương của gradient trước đó, giúp tránh việc quá chậm khi học các tham số có gradient nhỏ và giúp nhanh chóng hội tụ cho các tham số có gradient lớn.

- Backpropagation with Weight Regularization:

Trong các bài toán học sâu, để tránh hiện tượng quá khớp (overfitting), một số kỹ thuật như L1 hoặc L2 regularization có thể được áp dụng cùng với thuật toán Backpropagation. Điều này giúp giới hạn độ phức tạp của mô hình và ngăn chặn việc quá khớp dữ liệu huấn luyện.

Những biến thể này giúp thuật toán Backpropagation hoạt động hiệu quả hơn, đặc biệt trong các mô hình phức tạp và các tập dữ liệu lớn.

## Hệ thống gợi ý

### Giới thiệu về hệ thống gợi ý

Sự bùng nổ về lượng thông tin kỹ thuật số và số lượng người dùng internet trên toàn cầu đã đặt ra những thách thức lớn về tình trạng quá tải thông tin, gây khó khăn trong việc truy cập và sử dụng thông tin một cách hiệu quả. Để giải quyết vấn đề, các công ty hàng đầu như Netflix, Amazon, Shopee đã cách mạng hóa cách người dùng tiếp cận sản phẩm và dịch vụ, cho phép họ tận hưởng mọi tiện ích ngay tại nhà chỉ với vài cú nhấp chuột. Những nền tảng này ứng dụng thuật toán đề xuất tiên tiến, giúp cải thiện đáng kể trải nghiệm người dùng. Nhờ đó, người dùng được cung cấp các tùy chọn đa dạng, được thiết kế tinh tế dựa trên sở thích và nhu cầu cá nhân.Vì đáp ứng mong muốn và thị hiếu của người sử dụng, nhu cầu về các hệ thống gợi ý tăng hơn bao giờ hết.

Hệ thống gợi ý là hệ thống lọc thông tin giải quyết vấn đề quá tải thông tin bằng cách lọc ra đoạn thông tin quan trọng ra khỏi lượng lớn thông tin được tạo theo thị hiếu, sở thích hoặc hành vi quan sát của người dùng về sản phẩm. Như vậy có thể nói rằng hệ thống gợi ý được xây dựng nhằm đưa ra những gợi ý cho từng người dùng dựa trên thị hiếu của họ về sản phẩm. Hệ thống gợi ý có lợi cho cả nhà cung cấp dịch vụ (doanh nghiệp) và khách hàng sử dụng. Bằng cách này, các hệ thống gợi ý không chỉ làm tối ưu các chi phí và lựa chọn các sản phẩm trong môi trường kinh doanh online mà còn cải thiện đáng kể quá trình ra quyết định và hiệu quả. Như vậy, hệ thống gợi ý rất cần thiết không chỉ với các ứng dụng mà với việc quản lý và phân tích dữ liệu doanh nghiệp quan trọng hơn bao giờ hết.

Một hệ thống gợi ý hiệu quả thường dựa trên ba yếu tố cơ bản:

* Tập hợp người dùng (U):

Đây là danh sách chứa các thông tin cá nhân của người dùng đã đăng ký hoặc được theo dõi bởi hệ thống, ký hiệu là *U = {u1, u2, ..., up}*. Những thông tin này có thể bao gồm lịch sử mua hàng, sở thích, đánh giá sản phẩm, hoặc hành vi duyệt web.

* Tập hợp mục dữ liệu (I):

Tập hợp này chứa thông tin về các sản phẩm hoặc dịch vụ có sẵn trên hệ thống, thường được đánh dấu bằng các định danh và kèm theo các thuộc tính đặc trưng, được ký hiệu là *I = {i1, i2, ..., iq}*.

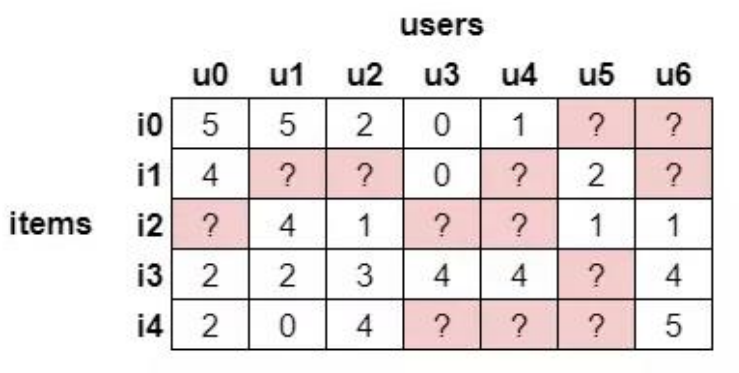
* Tập hợp mối quan hệ (R):

Đây là tập hợp các "mối quan hệ" giữa người dùng và mục dữ liệu, được biểu diễn dưới dạng các cặp *(ui, ij)* thể hiện người dùng ui có quan hệ như thế nào với mục dữ liệu *ij*. Có thể là qua việc mua hàng, xem sản phẩm, đánh giá hoặc tương tác khác.

Hiện nay, phương pháp phản hồi rõ ràng (explicit feedback) và phương pháp phản hồi ngầm (implicit feedback) là hai phương pháp phổ biến nhất để thu thập dữ liệu hồ sơ người dùng. Để tạo hồ sơ người dùng, hệ thống sẽ yêu cầu người dùng cung cấp xếp hạng cụ thể cho từng mục dữ liệu nếu phương pháp phản hồi rõ ràng. Dữ liệu khách hàng được truyền trực tiếp lên hệ thống tư vấn bằng cách thực hiện các bước chuyển đổi trung gian, và kết quả tư vấn được coi là đáng tin cậy nhất theo phương pháp này.

Mặc dù phương pháp này đưa ra kết quả đề xuất có độ tin cậy tốt hơn, nhưng nó yêu cầu công việc bổ sung trên hệ thống và có thể dẫn đến trải nghiệm kém hơn cho khách hàng. Ngoài ra, việc sử dụng các kỹ thuật thu thập phản hồi trong thực tế có thể sẽ gặp nhiều khó khăn vì khách hàng thường không muốn chia sẻ nhiều thông tin cá nhân.

Phương pháp thu thập phản hồi ẩn xử lý thông tin về sở thích của khách hàng bằng cách ghi nhận các thông tin truy vết mà người dùng tạo ra trên hệ thống, chẳng hạn như lịch sử mua hàng, thời gian xem trang web, số lần click chuột vào các siêu liên kết và lịch sử truy cập trang web. Điều này tạo điều kiện thuận lợi để tìm kiếm và khám phá thông tin về sở thích của khách hàng. Do yêu cầu hệ thống cần phải phải tiến hành các thao tác chuyển đổi ban đầu để thu thập dữ liệu về hành vi của người tiêu dùng, phương pháp phản hồi tường minh có khả năng nâng cao trải nghiệm người tiêu dùng với hệ thống.

Ma trận tương tác User—Item là một tập hợp dữ liệu được lưu trữ cho thấy sở thích của mỗi người dùng đối với các sản phẩm trong hệ thống. Dữ liệu này có thể được hiển thị và lưu trữ trong cơ sở dữ liệu dưới dạng ma trận, trong đó mỗi hàng tương ứng với một người dùng và mỗi cột tương ứng với một sản phẩm. Giá trị đánh giá của người dùng đối với mặt hàng được hiển thị trong từng ô của ma trận chính.

Hình 2‑5. Ma trận tương tác User - Item

Khách hàng sẽ thường không đánh giá tất cả những gì họ đã trải qua vì lý do tâm lý. Họ thường chỉ đưa ra đánh giá các mục với tình trạng tích cực, tức là rất thích hoặc tiêu cực. Do đó, ma trận tương tác người dùng—sản phẩm nói trên thường thiếu giá trị ở nhiều ô. Đây được coi là vấn đề về sự không đầy đủ của dữ liệu hoặc dữ liệu thưa. Các phương pháp gợi ý dựa trên bộ nhớ, còn được gọi là các thuật toán gợi ý dựa trên bộ nhớ, gặp khó khăn khi dữ liệu của ma trận tương tác bị thưa. Ngoài ra, các hàng hoặc cột trong ma trận Khách hàng–Sản phẩm sẽ trống hoàn toàn khi có tương tác giữa khách hàng hoặc mặt hàng mới trong hệ thống. Trong các hệ thống gợi ý dựa trên cộng tác, Hiện tượng này được biết đến với tên gọi là vấn đề khởi động nguội. Khả năng tạo gợi ý bằng cách kết hợp các dữ liệu dựa trên sở thích tương tự của các nhóm người dùng và mục bị hạn chế khi hệ thống đối mặt với các mục không có tương tác.

### Các phương pháp tiếp cận khi xây dựng hệ thống gợi ý

#### Lọc nội dung

Gợi ý dựa trên nội dung (Content based filtering – CBF) là một thuật toán phụ thuộc vào miền và nó nhấn mạnh nhiều hơn vào việc phân tích các thuộc tính của các mục để tạo ra dự đoán. Khi các tài liệu như trang web, ấn phẩm và tin tức được đề xuất, kỹ thuật lọc dựa trên nội dung là kỹ thuật lọc dựa trên nội dung thành công nhất dựa trên nội dung, gợi ý được thực hiện dựa trên hồ sơ người dùng bằng cách sử dụng các tính năng được trích xuất từ nội dung của các mục mà người dùng đã đánh giá trong quá khứ.

#### Lọc cộng tác

Lọc cộng tác (Collaborative filtering – CF) là một kỹ thuật dự đoán độc lập với miền cho nội dung không thể được mô tả dễ dàng và đầy đủ bằng siêu dữ liệu như phim và nhạc. Kỹ thuật lọc cộng tác hoạt động bằng cách xây dựng cơ sở dữ liệu (ma trận mục người dùng) về các tùy chọn cho các mục của người dùng. Sau đó, nó khớp người dùng với sở thích và sở thích có liên quan bằng cách tính toán sự tương đồng giữa các hồ sơ của họ để đưa ra xuất.

Quy trình đề xuất dựa trên sự hợp tác như sau:

* Input:

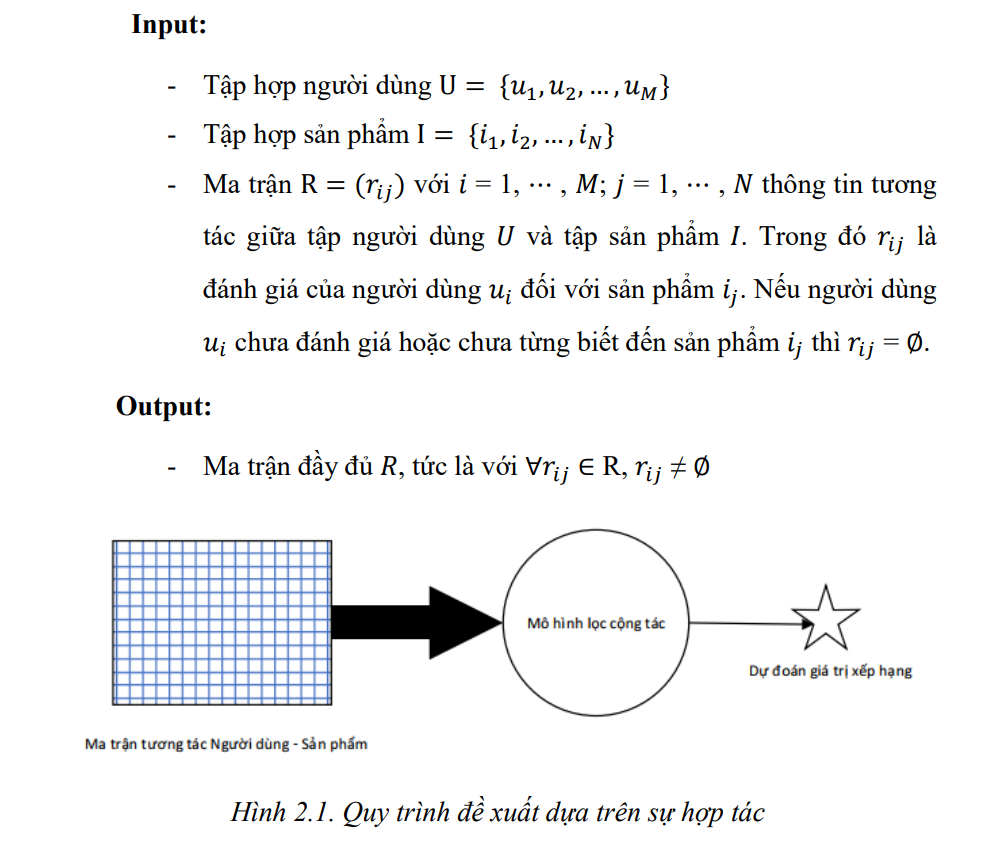
+ Tập hợp người dùng *U = {u1, u2, ..., uM}*

+ Tập hợp sản phẩm *I = {i1, i2, ..., iN}*

+ Ma trận *R = (rij)* với *i =* ; *j =*  thông tin tương tác giữa tập người dùng *U* và tập sản phẩm *I*. Trong đó *rij* là đánh giá của người dùng ui đối với sản phẩm *ij*. Nếu người dùng ui chưa đánh giá hoặc chưa từng biết đến sản phẩm *ij*thì *rij* = .

* Output:

Ma trận đầy đủ R, tức là với  *rij* R, *rij*

 Dựa trên cơ sở đó, trong bài báo cáo này, nhóm chúng em tập trung vào việc sử dụng hệ thống lọc cộng tác để đưa ra gợi ý cho người tiêu dùng sản phẩm. Một trong những phương pháp phổ biến trong hệ thống lọc cộng tác là lọc cộng tác dựa trên mô hình (Model-Based Collaborative Filtering).

Hình 2‑6. Quy trình cơ bản của hệ thống lọc cộng tác

Lọc cộng tác dựa trên mô hình sử dụng học máy và thuật toán khai thác dữ liệu để tính trước và dự đoán. Sau đó, chiến lược này sử dụng các mô hình này để tính toán trực tiếp các giá trị xếp hạng của người dùng. Lọc cộng tác dựa trên mô hình giảm đáng kể chi phí tính toán và yêu cầu bộ nhớ trong khi vẫn đạt được độ chính xác tương đương với các phương pháp dựa trên bộ nhớ bằng cách sử dụng các phương pháp học máy hiện đại. Hơn nữa, khi dữ liệu thưa thớt xảy ra, phương pháp này cũng có thể khắc phục được những hạn chế của phương pháp dựa trên bộ nhớ.

Một số kỹ thuật lọc cộng tác phổ biến dựa trên mô hình:

* Luật kết hợp

Thuật toán luật kết hợp là một phương pháp trong khai thác dữ liệu, được sử dụng để xác định các mối quan hệ giữa các vật phẩm trong một tập dữ liệu lớn. Cụ thể, phương pháp này phân tích các giao dịch để tìm ra những quy tắc có thể dự đoán sự xuất hiện của một mục (item B) dựa trên sự xuất hiện của một hoặc nhiều mục khác (item A) trong cùng giao dịch. Một quy tắc kết hợp thường được biểu diễn dưới dạng A → B, nghĩa là khi A xuất hiện, B cũng có xu hướng xuất hiện.Phương pháp này hỗ trợ tạo ra các mô hình dự đoán từ dữ liệu giao dịch. Ví dụ, nếu khách hàng thường mua sữa và bánh mì cùng nhau, thuật toán luật kết hợp sẽ xác định mối liên hệ này và có thể sử dụng để đề xuất sản phẩm phù hợp trong tương lai. Ngoài ra, các thuật toán khai thác luật kết hợp còn giúp giảm kích thước mô hình so với dữ liệu ban đầu, từ đó nâng cao hiệu quả tính toán và tiết kiệm bộ nhớ.Đây là một phương pháp đã được chứng minh mang lại hiệu quả cao trong việc phân tích và khám phá dữ liệu, đặc biệt trong các ứng dụng đề xuất sản phẩm hoặc dịch vụ dựa trên sở thích của người dùng.

* Phân cụm

Phân cụm là một kỹ thuật quan trọng được ứng dụng trong nhiều lĩnh vực như nhận dạng mẫu, xử lý hình ảnh, phân tích dữ liệu thống kê, và khám phá tri thức. Kỹ thuật này chia một tập dữ liệu lớn thành các nhóm con, gọi là cụm, dựa trên độ tương đồng giữa các đối tượng. Mục tiêu là tạo ra các cụm với sự tương đồng cao bên trong và sự khác biệt lớn giữa các cụm khác nhau. Trong các hệ thống gợi ý, phân cụm có thể được sử dụng để nhóm các người dùng có sở thích tương tự. Điều này cho phép hệ thống dự đoán sở thích của một người dùng dựa trên đánh giá trung bình của nhóm mà người dùng đó thuộc về. Nếu một người dùng nằm trong nhiều nhóm khác nhau, hệ thống có thể đưa ra các đề xuất dựa trên giao thoa sở thích giữa các nhóm. K-means là một trong những thuật toán phân cụm phổ biến nhất, phân chia n đối tượng thành K cụm dựa trên các đặc điểm của dữ liệu. Đây là một phương pháp học không giám sát, tức là không yêu cầu nhãn dữ liệu trước khi thực hiện. Ngoài ra, bản đồ tự tổ chức (Self-Organizing Map - SOM) là một kỹ thuật phân cụm khác, dựa trên mạng nơ-ron nhân tạo và học không giám sát. SOM chuyển đổi không gian đa chiều của dữ liệu thành không gian hai chiều, trong khi vẫn giữ nguyên cấu trúc topo ban đầu của dữ liệu. Trong các thuật toán dựa trên hợp tác, phân cụm giúp giảm số lượng cụm cần xem xét, làm tăng hiệu quả tính toán. Kỹ thuật này giảm kích thước của ma trận người dùng-mục tiêu cần xử lý, từ đó nâng cao hiệu suất của hệ thống gợi ý.

* Cây quyết định

Là một công cụ mạnh mẽ và dễ hiểu trong lĩnh vực học máy, được sử dụng để thực hiện các nhiệm vụ phân loại và hồi quy. Chúng hoạt động bằng cách phân tích một tập hợp dữ liệu huấn luyện đã được gắn nhãn, từ đó xây dựng một mô hình dạng cây với các nút đại diện cho các quyết định dựa trên thuộc tính của dữ liệu và các lá của cây đại diện cho kết quả phân loại hoặc giá trị đầu ra. Thuật toán cây quyết định sử dụng mô hình cây này để phân loại hoặc dự đoán đầu ra cho các mẫu dữ liệu mới không có trong tập huấn luyện. Một trong những ưu điểm chính của cây quyết định là tính trực quan và dễ hiểu; chúng cho phép chúng ta dễ dàng theo dõi cách một quyết định được đưa ra thông qua một loạt các câu hỏi và quyết định dựa trên các thuộc tính dữ liệu. So với các thuật toán phức tạp khác như Mạng nơ-ron nhân tạo (ANN) và Máy vector hỗ trợ (SVM), cây quyết định thường được coi là dễ hiểu hơn nhiều bởi vì chúng thể hiện các quyết định và phân loại dữ liệu một cách rõ ràng và logic. Một điểm mạnh khác của cây quyết định là khả năng xử lý dữ liệu hỗn hợp – tức là, chúng có thể đồng thời xử lý dữ liệu số (liên tục) và danh mục (phân loại) mà không cần phải chuyển đổi hay tiền xử lý dữ liệu một cách phức tạp. Thậm chí, chúng còn có khả năng xử lý dữ liệu thiếu một cách hiệu quả, một tính năng quan trọng trong các tình huống mà dữ liệu thực tế thường không hoàn chỉnh. Cây quyết định có thể được sử dụng trong rất nhiều ứng dụng từ phân tích tín dụng, dự đoán khách hàng tiềm năng, đến chẩn đoán y tế và nhiều ngành khác, làm cho nó trở thành một công cụ đa năng trong học máy và phân tích dữ liệu

* Mạng nơron nhân tạo (ANN)

Mạng nơron là một cấu trúc gồm nhiều nút thần kinh được kết nối thành các lớp. Các trọng số được sử dụng để xác định mức độ ảnh hưởng của mỗi nút đối với các nút lân cận của nó. Mạng nơ-ron có thể mô phỏng các mối quan hệ phức tạp trong dữ liệu và giống như các hàm phi tuyến. Tuy nhiên, nhược điểm của mạng nơron là hiệu suất của chúng phụ thuộc rất nhiều vào cấu trúc của mạng, và là một thách thức để tìm ra cấu trúc mạng tối ưu.

#### Lọc kết hợp

Lọc kết hợp (Hybrid Filtering – HF) là một sự kết hợp giữa lọc cộng tác và lọc dựa trên nội dung, được giải thích qua ba giai đoạn. Giai đoạn đầu: hệ thống xác định người dùng có sự tương đồng với nhau bằng cách khớp hồ sơ người dùng. Giai đoạn thứ hai: hệ thống sẽ chọn các mục cho từng người dùng giống nhau bằng cách lấy các Vectơ tương ứng với hồ sơ người dùng và nội dung của mục đó. Giai đoạn cuối cùng: sau khi dự đoán cho mỗi mục bằng phương trình dự đoán, các mục sẽ được đề xuất cho người dùng mục tiêu. Phản hồi trực tiếp và gián tiếp.

### Neural Collaborative Filtering

#### Đặt vấn đề

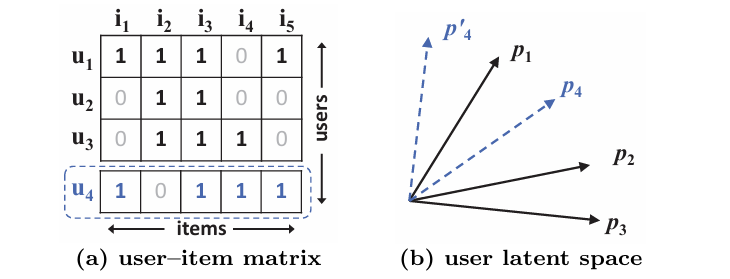
Giả sử M và N lần lượt biểu thị số lượng người dùng và sản phẩm. Chúng ta định nghĩa ma trận tương tác người dùng - sản phẩm *Y ∈ RM x N*  dựa trên phản hồi ngầm của người dùng như sau:

*yui = 1* nếu tương tác giữa người dùng u và sản phẩm mới được quan sát thấy.

*yui = 0* nếu tương tác giữa người dùng u và sản phẩm mới không được quan sát thấy.

Ở đây, giá trị yui = 1 chỉ ra rằng có một tương tác giữa người dùng u và sản phẩm i; tuy nhiên, điều này không nhất thiết có nghĩa là u thực sự thích i. Tương tự, yui = 0 không hẳn có nghĩa là u không thích i, mà có thể là người dùng chưa biết đến sản phẩm.

Vấn đề gợi ý dựa trên phản hồi ngầm được định nghĩa là bài toán ước lượng các điểm số của các mục không được quan sát trong y, điểm số này sẽ được sử dụng để xếp hạng các sản phẩm. Các phương pháp dựa trên mô hình giả định rằng dữ liệu có thể được tạo ra bằng mô hình cơ bản.

Cụ thể, chúng có thể được trừu tượng hóa như việc học , trong đó biểu thị điểm số dự đoán của tương tác. Để ước lượng các tham số , các cách tiếp cận hiện có thường tối ưu hóa một hàm mục tiêu. Hai loại hàm mục tiêu phổ biến nhất trong tài liệu là mất mát điểm (pointwise loss) và mất mát cặp (pairwise loss).

Hình 2‑7. Hạn chế của MF

Để giải quyết bài toán gợi ý dựa trên phản hồi ngầm, các phương pháp hiện đại thường tìm cách khai thác các đặc trưng ẩn trong dữ liệu nhằm dự đoán điểm số tương tác chưa được quan sát. Một cách tiếp cận phổ biến và hiệu quả trong lĩnh vực này là phân rã ma trận (Matrix Factorization), nơi mỗi người dùng và mỗi sản phẩm được biểu diễn thông qua các vector đặc trưng ẩn.

Phân rã ma trận (MF) gán cho mỗi người dùng và mỗi sản phẩm một vector giá trị thực chứa các đặc trưng ẩn. Gọi *Pu*và *Qi* lần lượt là vector ẩn của người dùng u và sản phẩm i. MF ước tính một tương tác *yui* dưới dạng tích vô hướng của *Pu*và *Qi*:

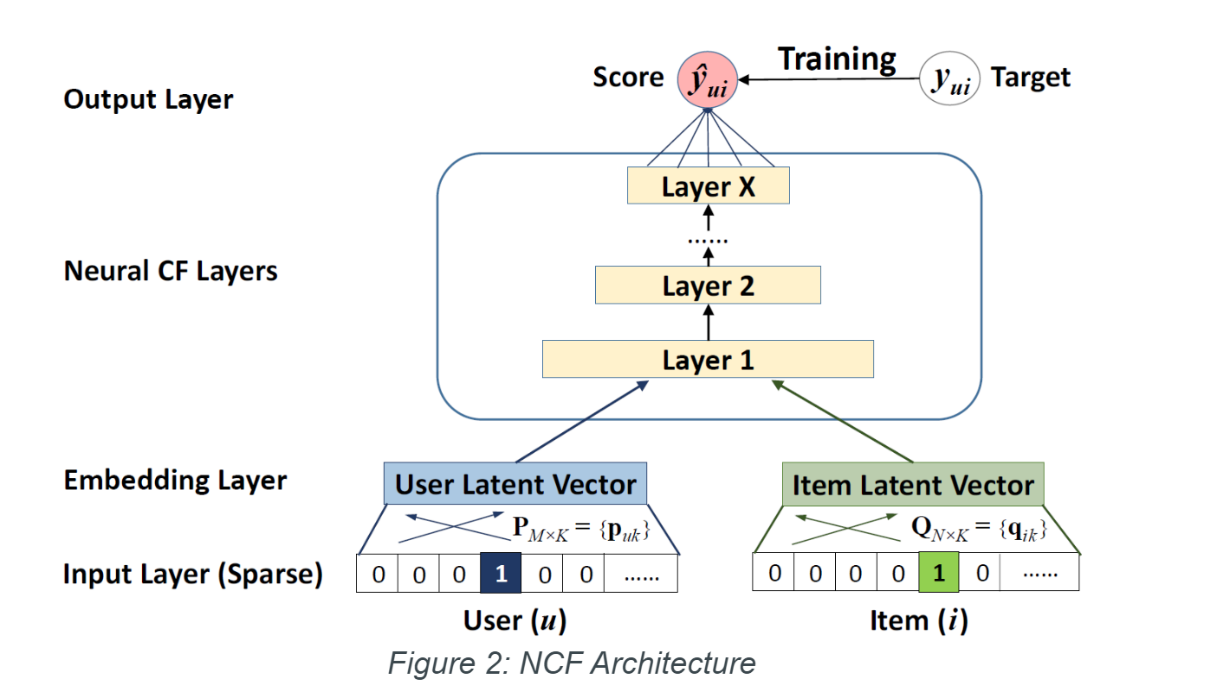
Trong đó: K biểu thị số chiều không gian đặc trưng ẩn.

MF giả định rằng mỗi chiều của không gian ẩn là độc lập với nhau và kết hợp chúng theo cách tuyến tính với cùng một trọng số. Vì vậy, MF có thể được xem là một mô hình tuyến tính của các đặc trưng ẩn.

Tuy nhiên,hạn chế của MF nằm ở việc sử dụng hàm tương tác tuyến tính đơn giản như tích vô hướng, điều này có thể không đủ để nắm bắt các cấu trúc phức tạp trong dữ liệu tương tác của người dùng.

#### Kiến trúc NCF

Để cho phép áp dụng hoàn toàn mạng nơ-ron vào lọc cộng tác, ta áp dụng một biểu diễn nhiều tầng để mô hình hóa một tương tác người dùng – sản phẩm *yui*. Ở đây, đầu ra của một tầng sẽ được sử dụng làm đầu vào cho tầng tiếp theo.

* Lớp đầu vào (Input Layer)

Hình 2‑8. Kiến trúc NCF

Lớp này bao gồm hai vector đặc trưng và mô tả người dùng u và sản phẩm i. Các vector này có thể được tùy chỉnh để hỗ trợ nhiều cách biểu diễn khác nhau của người dùng và sản phẩm, chẳng hạn như theo ngữ cảnh, dựa trên nội dung,... Trong báo cáo này, danh tính của người dùng và sản phẩm được sử dụng làm đặc trưng đầu vào ban đầu và được mã hóa dưới dạng vector nhị phân rải rác (one-hot encoding).

* Lớp nhúng (Embedding Layer)

Để làm cho các lớp thưa thớt dày đặc hơn, một lớp kết nối đầy đủ sẽ hoạt động như một lớp nhúng. Các vector này có thể được xem như vector đặc trưng tiềm ẩn của người dùng và sản phẩm để làm đầu vào của lớp cộng tác nơron.

* Lớp lọc cộng tác nơron (Neural CF Layers)

Vector nhúng của người dùng và sản phẩm được đưa qua một kiến trúc mạng nơ-ron nhiều tầng để ánh xạ chúng thành các điểm số dự đoán. Mỗi tầng trong mạng nơ-ron này có thể được tùy chỉnh để khám phá các cấu trúc tiềm ẩn nhất định trong tương tác giữa người dùng và sản phẩm.

+ Hàm tương tác của NCF tính bằng công thức:

Trong đó:

và lần lượt biểu thị hàm ánh xạ cho tầng đầu ra và lớp x trong mạng CF nơron, và có tổng cộng X lớp CF nơ-ron.

+ Hàm mất mát bình phương của NCF được tính theo công thức:

Hàm mất mát bình phương của NCF, trong đó biểu thị tập hợp các tương tác được quan sát trong ma trận phản hồi ngầm, và biểu thị tập hợp các trường hợp âm tính, có thể là tất cả hoặc được lấy mẫu từ các tương tác không được quan sát; và là một siêu tham số biểu thị trọng số của trường hợp đào tạo (u, i).

* Lớp đầu ra (Output Layer)

Lớp này đưa ra điểm số dự đoán và mô hình được huấn luyện bằng cách tối thiểu hóa mất mát giữa và giá trị thực . Để tối thiểu hóa ta dùng hàm sau:

Đây là hàm mục tiêu cần được tối thiểu hóa cho các phương pháp NCF, và việc tối ưu hóa có thể được thực hiện bằng cách thực hiện gradient descent ngẫu nhiên (SGD). Nhận thấy rằng nó giống với hàm mất mát Binary Cross-Entropy, còn được gọi là Log loss.

#### Phân tích ma trận tổng quát hóa (Generalized Matrix Factorization - GMF)

Vì MF là thuật toán được sử dụng rộng rãi cho việc đề xuất, việc chứng minh rằng NCF là trường hợp tổng quát của MF cho phép NCF bao trùm một nhóm lớn các mô hình nhân tố hóa. Khi các vectơ mục và người dùng được mã hóa một lần và được nhúng để tạo thành các vectơ dày đặc nên chúng hoạt động tương tự như các vectơ đặc trưng tiềm ẩn được sử dụng trong MF. Đặt vectơ tiềm ẩn của người dùng *pu* là và vectơ tiềm ẩn của mục *qi* là . Chúng ta định nghĩa hàm ánh xạ 1 của lớp NCF đầu tiên bằng công thức:

Dự đoán cuối cùng được hiển thị trong hình. Nếu chúng ta sử dụng hàm nhận dạng cho *aout* và *h* là một vectơ trọng số đơn vị, thì nó giống với phương trình MF cho dự đoán mục tiêu. Do đó chứng minh rằng MF là một trường hợp đặc biệt của NCF.

Ta đã chứng minh rằng MF là trường hợp đặc biệt của NCF, chúng ta có thể thu được MF tổng quát và mở rộng từ công thức dự đoán NCF (công thức trên cùng của Phương trình 8). h vectơ trọng số được học từ dữ liệu thay vì cố định nó dưới dạng hằng số, mô hình kết quả sẽ là một biến thể của MF xem xét tầm quan trọng khác nhau của các tính năng tiềm ẩn. Mặt khác, nếu aout hàm kích hoạt của lớp cuối cùng được lấy là hàm phi tuyến tính thì bộ đề xuất sẽ có tính biểu cảm cao hơn so với mô hình tuyến tính. Để tạo phần Hệ số ma trận tổng quát (GMF), một phiên bản tổng quát của MF sử dụng hàm kích hoạt sigmoid và trọng số học từ dữ liệu sẽ được triển khai.

#### Multi- Layer Perceptron (MLP)

Hình 2‑9. Kiến trúc GMF

Do NCF áp dụng hai luồng mô hình hóa cho người dùng và sản phẩm, việc kết hợp các đặc trưng của hai luồng này là điều tự nhiên. Một thiết kế phổ biến là nối (concatenate) các vector đặc trưng này lại với nhau, như đã được áp dụng rộng rãi trong các công trình học sâu đa chiều (multimodal deep learning).

Tuy nhiên, việc nối đơn thuần các vector đặc trưng không tính đến bất kỳ tương tác nào giữa đặc trưng người dùng và sản phẩm, điều này là không đủ để mô hình hóa hiệu ứng lọc cộng tác. Để giải quyết vấn đề này, ta thêm các tầng ẩn vào vector nối bằng cách sử dụng mạng perceptron nhiều tầng (MLP) nhằm học các tương tác giữa các đặc trưng ẩn của người dùng và sản phẩm.

Mô hình MLP trong khung NCF được định nghĩa như sau:

Trong đó:

Wx​  - Ma trận trọng số

bx​ – Vector bù trừ

a​x – hàm kích hoạt của tầng x.

#### Mô hình phân rã ma trận nơron (Neural Matrix Factorization – NeuMF)

Nơ-ron lọc cộng tác (NCF) được sử dụng cho bài toán đề xuất sử dụng phản hồi gián tiếp. Phản hồi gián tiếp có mặt khắp mọi nơi trong các hệ thống đề xuất. Các hành động như nhấn chọn, mua và xem là những phản hồi gián tiếp phổ biến có thể dễ dàng thu thập và thể hiện được sở thích của người tiêu dùng. NeuMF, hướng tới việc giải quyết các vụ xếp hạng cá nhân hóa sử dụng phản hồi gián tiếp. Mô hình này tận dụng tính linh hoạt và tính phi tuyến của mạng nơ-ron để thay thế tích vô hướng trong phân rã ma trận, nhằm nâng cao tính biểu diễn của mô hình. Cụ thể, mô hình này gồm hai mạng con là phân rã ma trận tổng quát GMF và MLP, và mô hình hóa các tương tác theo hai mạng này thay vì các tích vô hướng đơn giản. Kết quả đầu ra của hai mạng này được ghép nối với nhau để tính điểm dự đoán cuối cùng, mô hình này sinh ra danh sách đề xuất đã được xếp hạng cho từng người tiêu dùng dựa trên phản hồi gián tiếp.

* Bước 1:

Sử dụng cùng một lớp nhúng cho cả hai cấu trúc và kết hợp đầu ra.

GMF và MLP phải sử dụng cùng một kích thước lớp nhúng; đối với các tập dữ liệu mà kích thước embedding tối ưu của hai mô hình khác nhau đáng kể, điều này làm hạn chế hiệu suất của mô hình.

Tính toán đầu ra với cùng lớp nhúng trong mô hình kết hợp

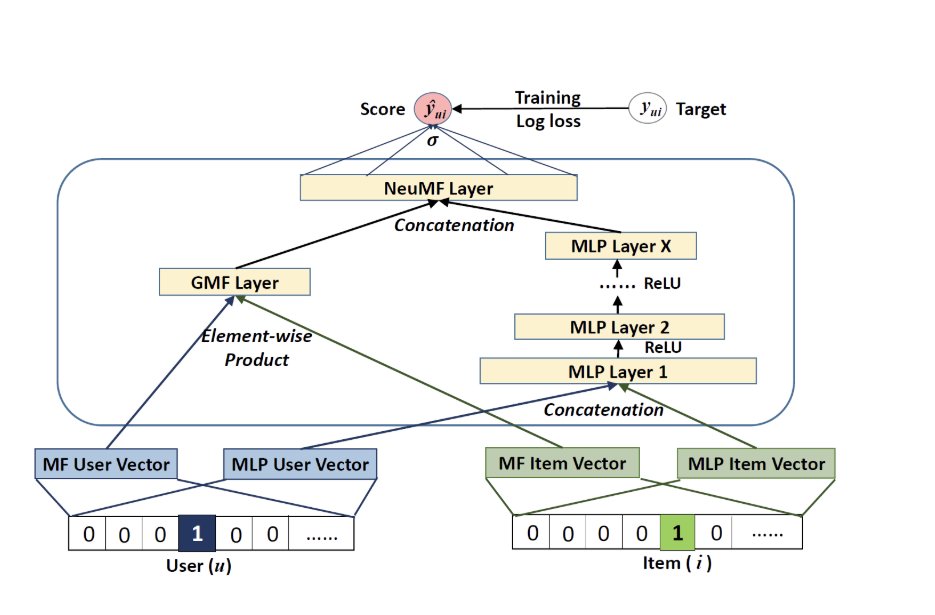
Trong đó: h, W, b, và a lần lượt biểu thị vector trọng số của GMF, ma trận trọng số của MLP, vector độ lệch (bias), và hàm kích hoạt.

* Bước 2:

Sử dụng các lớp nhúng riêng biệt cho cả hai cấu trúc và kết hợp đầu ra.

Lớp nhúng cho các vector người dùng và sản phẩm được đưa riêng biệt vào mô hình MLP và mô hình GMF. Các đầu ra thu được từ cả hai mô hình được nối lại tại lớp ẩn cuối cùng và tạo thành mô hình hoàn chỉnh NeuMF.

Tính toán đầu ra bằng công thức:

Trong đó:

Hình 2‑10. Kiến trúc NeuMF

và tương ứng biểu thị nhúng người dùng cho phần GMF và MLP; và ký hiệu tương tự và cho nhúng sản phẩm.

### Phương pháp và chỉ số đánh giá hệ thống gợi ý

* Leave One Out Validation (LOOV)

Leave One Out Validation (kiểm chứng bỏ qua một dữ liệu) là một phương pháp phổ biến được sử dụng để đánh giá các hệ thống gợi ý, trong đó một điểm dữ liệu duy nhất được loại bỏ khỏi tập dữ liệu và mô hình được đào tạo trên các điểm dữ liệu còn lại. Điểm dữ liệu đã xóa sau đó được sử dụng làm tập kiểm tra để đánh giá hiệu suất của mô hình. Quá trình này được lặp lại cho từng điểm dữ liệu trong tập dữ liệu. Trong chủ đề này, đối với mỗi khách hàng, một giao dịch gần nhất (hoặc một mục tương tác cuối cùng) được chọn làm tập kiểm tra và các giao dịch còn lại được sử dụng làm tập huấn luyện. Mục tiêu là kiểm tra khả năng của mô hình trong việc dự đoán chính xác mục mà khách hàng có khả năng tương tác tiếp theo dựa trên dữ liệu lịch sử.

* Chỉ số đánh giá:

+ Hit Ratio(HR)

Hit ratio (tỷ lệ trúng) là một chỉ số được sử dụng trong hệ thống gợi ý để đo lường hiệu quả của hệ thống trong việc đưa ra các đề xuất chính xác. Chỉ số này đo lường tỷ lệ người dùng nhận được ít nhất một khuyến nghị có liên quan. Đối với chủ đề này, Hit Ratio được ký hiệu là HR@K (K - top K sản phẩm được gợi ý cho khách hàng) cho biết tỷ lệ khách hàng có ít nhất một mục có liên quan trong K. Nếu khách hàng nào có ít nhất một mục liên quan sẽ được tính là một hit, ngược lại sẽ không được tính. Sau khi thực hiện trên toàn bộ tệp khách hàng, tỷ lệ HR@K sẽ được tính theo công thức:

Trong đó:

HR@K - Tỷ lệ trúng đích ở top K

No of Hits - Số lần trúng đích

No Of Users - Tổng số khách hàng

+ Normalized Discounted Cumulative Gain(NDCG)

Lợi nhuận tích lũy chiết khấu chuẩn hóa (NDCG) là một số liệu đánh giá chất lượng của các công cụ tìm kiếm, hệ thống đề xuất và các thuật toán xếp hạng khác. NDCG đo lường chất lượng của danh sách xếp hạng có tính đến mức độ liên quan của các mục và thứ tự xếp hạng (giá trị từ 0 đến 1). Đối với chủ đề này, NCDG được ký hiệu là NDCG@K (K - top K sản phẩm được xem xét đến) cho biết hiệu suất của một hệ thống xếp hạng khi chỉ xem xét K mục đầu tiên trong danh sách. Giá trị càng gần 1, hệ thống xếp hạng càng chính xác và hiệu quả trong việc đưa ra các mục quan trọng ở những vị trí đầu tiên trong danh sách. NDCG được tính theo công thức sau:

Trong đó:

Lợi nhuận tích lũy chiết khấu (Discounted Cumulative Gain - DCG): là phiên bản có trọng số của CG (tổng các điểm liên quan của các kết quả được xếp hạng trong danh sách trả về), sử dụng logarit để giảm relevance score tương ứng với vị trí của các kết quả.

Trong đó:

– mức độ liên quan được phân loại của kết quả tại vị trí i

Lợi nhuận tích lũy chiết khấu lý tưởng (Ideal Discounted Cumulative Gain - IDCG):

Trong đó:

– biểu thị danh sách các tài liệu có liên quan (được sắp xếp theo mức độ liên quan) trong ngữ liệu cho đến vị trí K.

# ỨNG DỤNG HỌC XÂY DỰNG HỆ THỐNG GỢI Ý SẢN PHẨM CHO NGƯỜI TIÊU DÙNG

## Tổng quan về hành vi người tiêu dùng

### Khái niệm

Theo Philip Kotler (2001), hành vi người tiêu dùng được định nghĩa là : “Một tổng thể những hành động diễn biến trong suốt quá trình kể từ khi nhận biết nhu cầu cho đến khi mua và sau khi mua sản phẩm”.

Bảng 3‑1. Mô hình hành vi người tiêu dùng

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Kích thích Marketing | Các yếu tố môi trường |  | Đặc điểm người mua | Quá trình ra quyết định mua |  | Các đáp ứng của người tiêu dùng |
| Sản phẩm  Giá cả  Vị trí  Chiêu thị | Kinh tế  Công nghệ  Pháp luật  Văn hóa |  | Văn hóa  Xã hội  Cá nhân  Tâm lý | Nhận thức vấn đề  Tìm kiếm thông tin  Đánh giá  Quyết định  Hành vi mua |  | Lựa chọn sản phẩm  Lựa chọn nhãn hiệu sản phẩm  Lựa chọn nơi mua  Định thời gian mua  Số lượng, tần suất mua |

Như vậy, có thể tóm gọn rằng hành vi người tiêu dùng (Consumer Behavior) là tập hợp các hành động, quyết định và quá trình tư duy của khách hàng trong việc tìm kiếm, lựa chọn sản phẩm hoặc sử dụng dịch vụ. Nói cách khác, hành vi người tiêu dùng chính là quá trình khách hàng đưa ra quyết định để chọn lựa thương hiệu và sử dụng sản phẩm hoặc dịch vụ.

### Vai trò nghiên cứu hành vi người tiêu dùng

- Phân khúc khách hàng:

Phân chia khách hàng thành các nhóm khác nhau như khách hàng trung thành, khách hàng mới, khách hàng sắp rời bỏ, hay khách hàng có giá trị cao, từ đó đưa ra chiến lược tiếp cận phù hợp cho từng nhóm.

- Tăng hiệu quả marketing:

Dựa vào phân tích , doanh nghiệp có thể triển khai các chiến dịch tiếp thị tập trung và cá nhân hóa, tối ưu hóa nguồn lực và tăng tỷ lệ phản hồi từ khách hàng.

- Dự đoán hành vi khách hàng:

Phân tích giúp dự đoán khả năng mua lại của khách hàng hoặc xác định nguy cơ rời bỏ, từ đó hỗ trợ các chiến lược giữ chân hiệu quả.

- Tối ưu hóa chi phí:

Bằng cách ưu tiên các nhóm khách hàng có tiềm năng cao, doanh nghiệp giảm được chi phí tiếp cận và chăm sóc không cần thiết đối với những khách hàng ít giá trị.

- Định hướng phát triển sản phẩm/dịch vụ:

Thông tin từ việc phân tích, cung cấp góc nhìn sâu sắc về nhu cầu và thói quen mua sắm, giúp doanh nghiệp cải thiện sản phẩm, dịch vụ để phù hợp hơn với từng phân khúc khách hàng.

## Tập hợp dữ liệu

Online Retail II(được lấy từ UCI Machine Learning Repository) chứa tất cả các giao dịch xảy ra đối với một cửa hàng bán lẻ trực tuyến từ ngày 01/12/2009 đến ngày 09/12/2011 có trụ sở tại Anh. Bộ dữ liệu này chứa 1,067,371 hàng và 8 cột bao gồm:

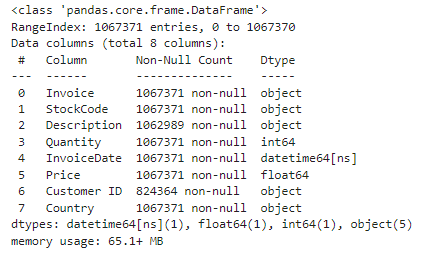
* Invoice: Mã hóa đơn gồm 6 chữ số được gán duy nhất cho mỗi giao dịch. Nếu mã bắt đầu bằng chữ ‘C’ thì nó cho biết đơn hàng bị hủy
* StockCode: Mã sản phẩm gồm 5 chữ số được gán duy nhất cho mỗi sản phẩm riêng biệt
* Description: Tên mặt hàng
* Quantity: Số lượng của từng sản phẩm trên mỗi giao dịch
* InvoiceDate: Ngày và giờ khi một giao dịch được tạo
* Price: Đơn giá. Đơn vị là bảng Anh
* CustomerID: Mã khách hàng gồm 5 chữ số được gán duy nhất cho mỗi khách hàng
* Country: Tên quốc gia nơi khách hàng cư trú

## Phân tích khai phá dữ liệu:

### Xử lý dữ liệu

* Tổng quan về dữ liệu:

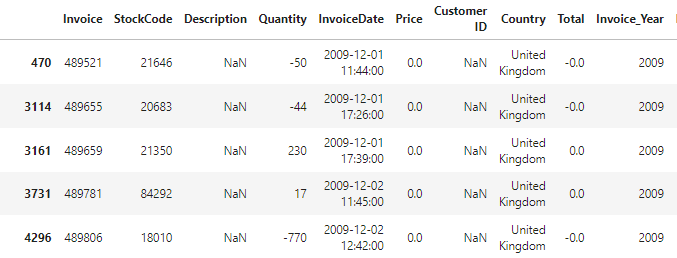
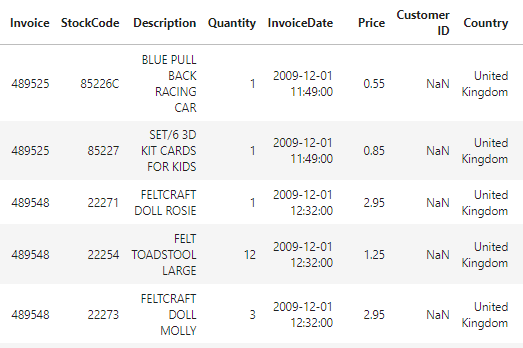
Dưới đây là 5 dòng ngẫu nhiên của bộ dữ liệu

Ta thấy cột Description có 4382 giá trị NaN và Customer ID có 243007 giá trị NaN. Đây là một số lượng NaN đáng kể trong bộ dữ liệu. 

Hình 3‑1. 5 dòng ngẫu nhiên của bộ dữ liệu

Hình 3‑2. Tóm tắt bộ dữ liệu

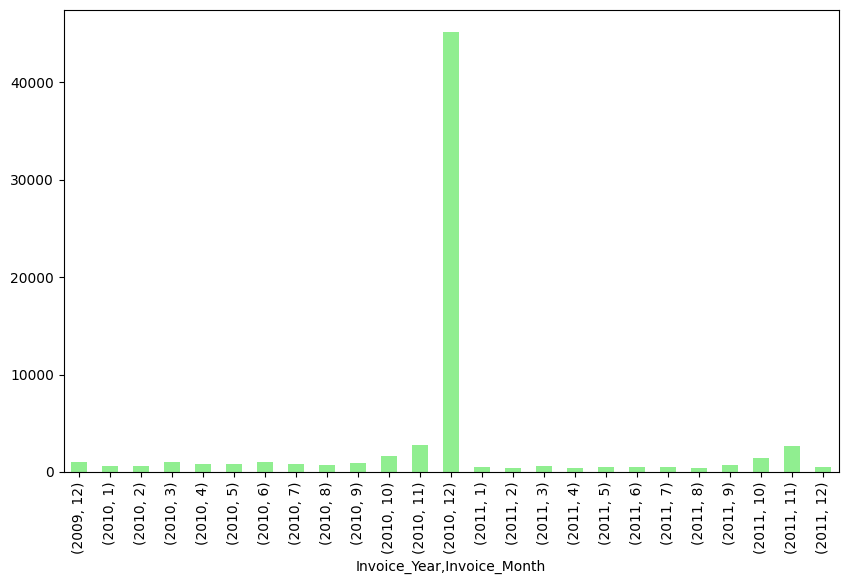
Trước tiên là cột Description, tất cả những dòng có giá trị NaN trong cột này thì giá bán điều bằng 0 và cũng không có mã khách hàng trong cột Customer ID. Có thể kết luận khả năng cao những đơn hàng đó bị hủy và giá bằng 0 cho thấy không có doanh thu từ những đơn này. Những giá trị NaN có thể là do khách hàng hủy đơn. Vì vậy, quyết định loại bỏ những giá trị NaN khỏi bộ dữ liệu.

Tiếp theo đến cột Customer ID, quan sát dữ liệu thấy rằng, những dòng NaN có phát sinh doanh thu. Vì thế có thể kết luận rằng, khách hàng có thể không cần đăng nhập tài khoản vẫn có thể mua hàng được. Nên nhóm quyết định không loại bỏ những giá trị NaN này. 

Hình 3‑3. 5 dòng đầu tiên có giá trị NaN trong cột Customer ID

Hình 3‑4. 5 dòng đầu tiên có giá trị NaN trong cột Description

Tiếp theo nhóm kiểm tra xem bộ dữ liệu có dữ liệu bị trùng lặp không. Kết quả là có 67026 dữ liệu trùng lặp.

Tuy nhiên sau khi quan sát kĩ các hàng trùng lặp thì thấy rằng các mặt hàng trùng lặp thường được liệt kê trên cùng một hóa đơn. Có thể đã xảy ra lỗi kỹ thuật khiến hệ thống không thể cập nhật số lượng chính xác và thay vào đó tạo các đơn hàng riêng biệt cho từng mặt hàng đã mua. Vì vậy nhóm quyết định không loại bỏ dòng trùng lặp khỏi dữ liệu vì chúng có thể quan trọng để hiểu cách khách hàng mua nhiều mặt hàng cùng nhau và xác định bức tranh chính xác hơn về hành vi mua hàng của khách hàng. 

Hình 3‑5. Biểu đồ số lượng dòng trùng lặp hàng tháng

Hình 3‑6. 5 dòng trungg lặp đầu tiên

Quan sát biểu đồ cột về số lượng dữ liệu trùng lặp qua từng tháng cho thấy sự tăng đột biến về số dòng trùng lặp trong tháng 12/2010 với khoảng 45000 dữ liệu trùng lặp. Điều này có thể có nghĩa là đã xảy ra sự cố với hệ thống trong tháng đó.

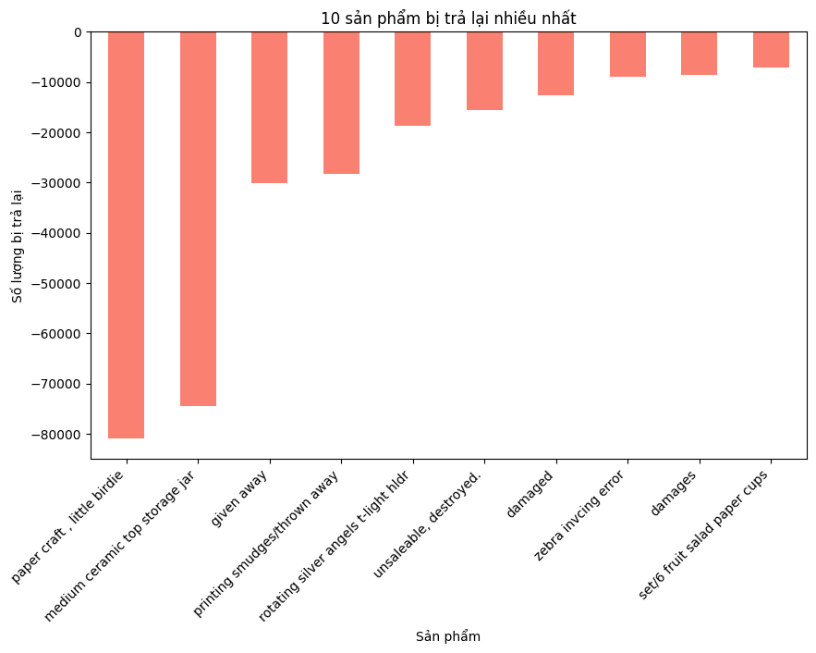
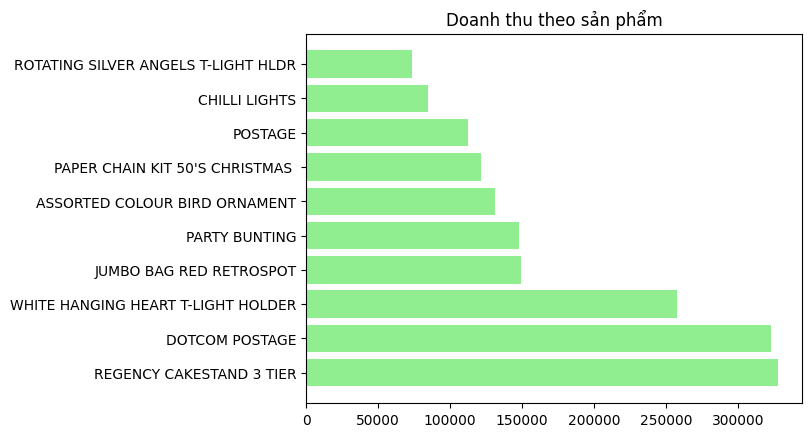
### Phân tích dữ liệu

#### Phân tích doanh thu theo sản phẩm

Hình 3‑7. Biểu đồ top 10 sản phẩm bán chạy nhất

Ta thấy rằng sản phẩm Tàu lượn thế chiến II là sản phẩm bán chạy nhất với hơn 100000 chiếc. Tuy nhiên, các sản phẩm tiện dụng cũng bán khá chạy. Các sản phẩm lọt top đều có điểm chung là dùng để phục vụ kinh doanh(vỏ đựng bánh, đồ ăn, đồ tặng kèm, trong đó mô hình máy bay này từng rất phổ biến trong việc là quà tặng kèm khi mua sản phẩm)

Có thể thấy top 10 đều là sản phẩm có thể gọi chung là hàng tiêu dùng hoặc hàng tiện dụng. Đây đều là sản phẩm được mua thường xuyên trong thời gian ngắn và không có giá trị lâu dài. Chúng thuộc danh mục FMCG vì chúng được tiêu thụ nhanh chóng và có nhu cầu mua lại cao.

Qua 2 biểu đồ ta thấy rằng mặc dù Tàu lượn thế chiến II là sản phẩm bán chạy nhất nhưng lại không nằm trong top 10 sản phẩm có doanh thu cao nhất. Tuy nhiên sản phẩm bán chạy thứ 2 nằm trong top 3 sản phẩm doanh thu cao nhất.

Hình 3‑8. Biểu đồ top 10 sản phẩm bị trả lại nhiều nhất

Hình 3‑9. Biểu đồ top 10 sản phẩm có doanh thu cao nhất

Từ biểu đồ, ta thấy paper craft, little bridie; medium ceramic top storage jar; rotaling silver angels t-light hldr; set 6 fruit salad paper cups đều có điểm chung là những sản phẩm dễ vỡ, hư hỏng do quá trình vận chuyển. Bên cạnh đó, các lỗi về in ấn, sản phẩm bị hỏng hóc, có vấn đề trong quá trình vận chuyển, nhầm lẫn trong hóa đơn/ mã sản phẩm cũng chiếm số lượng lớn trong số những đơn bị trả lại. Sau khi tính toán kết quả trả về cho biết:

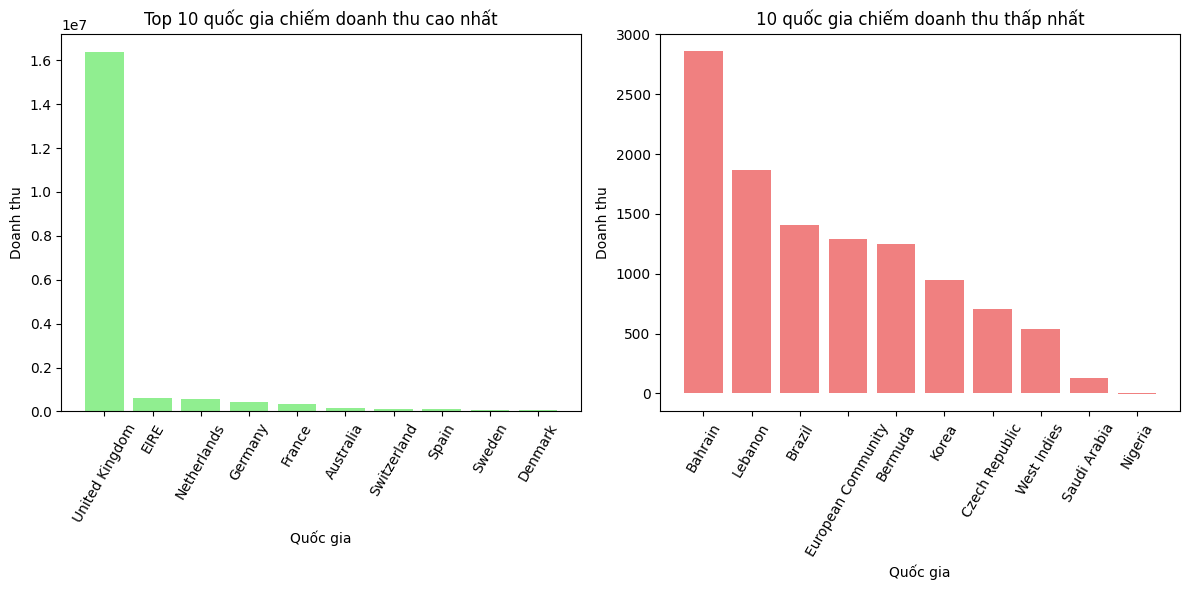
Tổng số lượng hàng mua: 11659609

Tổng số lượng hàng bị trả lại: 959955

Tỷ lệ hàng bị trả lại: 8.23%

Tỷ lệ trả hàng lên tới 8.23% cho thấy số lượng hàng trả lại không nhỏ, ảnh hướng tới lợi nhuận và uy tín của sàn. Từ biểu đồ ta cũng thấy rõ vấn đề chủ yếu nằm ở khâu vận chuyển, vì vậy công ty nên có sự cải thiện trong quy trình đóng gói vận chuyển chắn chắn và an toàn hơn.

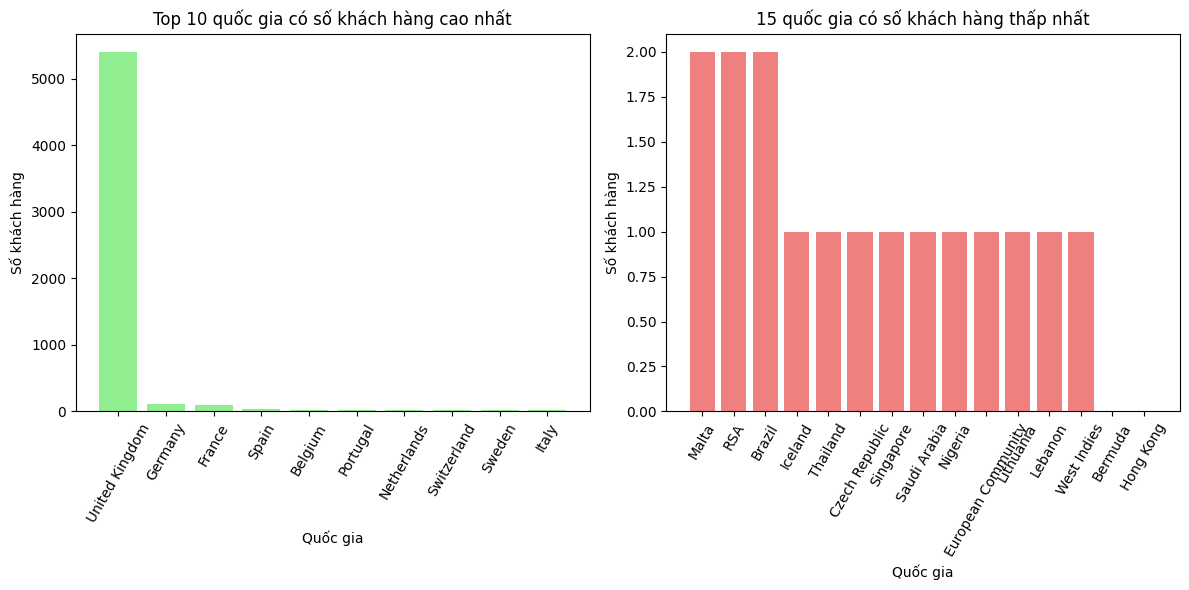
#### Phân tích doanh thu theo quốc gia

Nhìn vào top 10 quốc gia mang lại doanh thu cao nhất, Anh là quốc gia dẫn đầu với tổng lượng mua áp đảo và gấp khoảng 20 lần top 2 là EIRE. Top 10 quốc gia này đa phần nằm ở Châu Âu, chỉ trừ Úc.

Hình 3‑10. Biểu đồ cột top 10 quốc gia chiếm doanh thu cao nhất và thấp nhất

Nhìn vào top 10 quốc gia mang lại doanh thu thấp nhất, khách hàng từ Ả Rập Xê Út chỉ chi 131$ cho công ty trong 2 năm. Nói chung các quốc gia này đều không thuộc châu Âu, nhưng có 1 ngoại lệ là Cộng Hòa Séc.

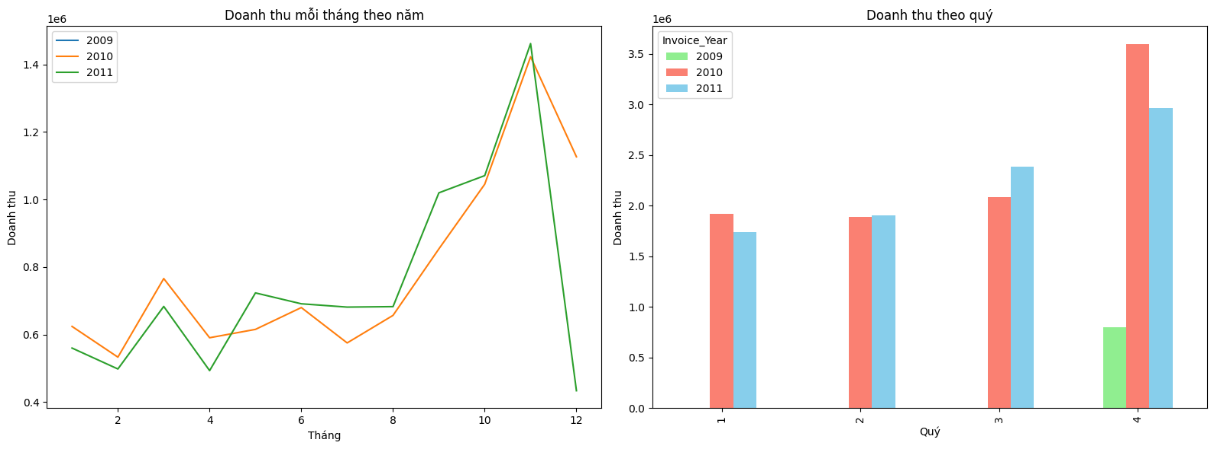
Để phân tích kỹ hơn, ta có thể chia thành 2 phần: Anh và phần còn lại.

Anh có khoảng 5400 khách hàng trong khi Đức và Pháp chỉ có 107 và 95 khách hàng. Sự khác biệt này cho thấy Anh thì thị trường chính của công ty. Ngoài ra có nhiều quốc gia chỉ có 1 khách hàng như Thái Lan, Nigeria... và Cộng Hòa Séc một lần nữa là một ngoại lệ trong danh sách này

Hình 3‑11. Biểu đồ top 10 quốc gia có số khách hàng cao nhất và thấp nhất

Qua 2 biểu đồ trên ta thấy được có sự tương quan giữa số lượng khách hàng và quốc gia.

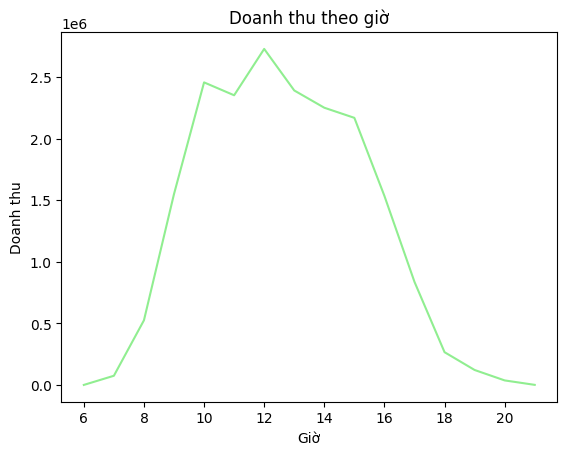
#### Phân tích doanh thu theo thời gian

Doanh thu tháng 2 và tháng 4 lần lượt giảm vào năm 2010 và 2011, nguyên nhân có thể là do hành vi giảm của khách hàng đối với một số sản phẩm nhất định, cũng có thể do hoạt động tiếp thị và khuyến mãi có nghĩa là nếu công ty thực hiện các chiến dịch tiếp thị hoặc khuyến mãi lớn trong những tháng cụ thể thì doanh thu có thể thấp hơn so với tháng trước đó. Hoạt động khuyến mãi giảm trong tháng 2 và tháng 4 có thể dẫn đến doanh thu thấp hơn

Hình 3‑12. Biểu đồ doanh thu theo tháng và quý

Những tháng đầu năm, ta có thể thấy doanh thu thấp hơn so với cuối năm, từ đặc điểm địa lý đã nói ở trên, đa phần các quốc gia là đối tượng chính đều là những nước có mùa đông lạnh, điều này ảnh hưởng đến thói quen tiêu dùng của khách hàng. Có thể lý giải rằng sau mùa lễ hội và mùa đông, người tiêu dùng thường có xu hướng tiết kiệm tiền và giảm chi tiêu. Thời tiết lạnh giá cũng ảnh hưởng đến du lịch, gián tiếp ảnh hưởng đến nhu cầu tiêu dùng những mặt hàng tiêu dùng không bền.

Vào những tháng cuối năm, khách hàng bắt đầu mua nhiều hơn để chuẩn bị cho các sự kiện cuối năm như giáng sinh, năm mới. Ta thấy tháng 11 của năm 2010 và 2011 đều tăng, ta có thể lý giải điều này dựa trên thói quen chuẩn bị cho giáng sinh và năm mới của Châu Âu và các nước đón Giáng sinh sẽ chuẩn bị trước đó cả 1 tháng.

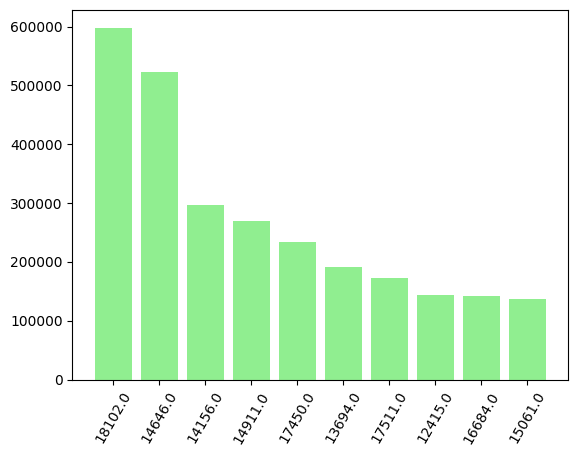
Nhìn chung người tiêu dùng thường dành nhiều ngân sách cho mua sắm quà tặng, tổ chức lễ hội vào các tháng 10, 11,12. Sau đó họ sẽ có xu hướng thắt chặt tài chính để tiết kiệm hoặc phục hồi dẫn đến doanh thu suy giảm. 

Hình 3‑13. Biểu đồ doanh thu theo giờ

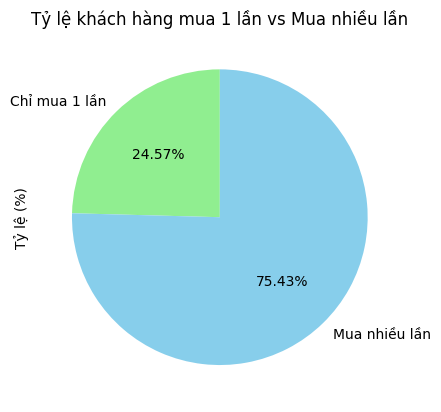
Từ biểu đồ này ta thấy doanh thu chủ yếu rơi vào khoảng thời gian từ 10h đến 13h và cao nhất là vào 12h. Đây là khung giờ vàng để công ty nên tập trung và tối ưu hóa hoạt động bán hàng hoặc tiếp thị trong những giờ có doanh thu cao này. Khung giờ 10-12h có doanh thu tăng cao, đây là khung giờ làm việc của đa số công ty châu Âu.

#### Phân tích khách hàng

Hình 3‑14. Biểu đồ top 10 khách hàng dành nhiều tiền để mua nhất

Khách hàng có ID là 18102 đã chi nhiều tiền nhất cho công ty. Khách hàng này đến từ Anh và đã trả tổng cộng 598000$. Khách hàng thứ 2 có ID là 14646 đến từ Hà Lan với 523000$. Nhìn chung thì top 10 khách hàng đến từ Anh, EIRE, Hà Lan và Úc đều là các quốc gia có nền kinh tế phát triển. Đây là những khách hàng VIP cần có các chiến lược chăm sóc khách hàng cho họ.

Dễ thấy đây đều là những đơn hàng lớn, khả năng rất cao đối tượng là doanh nghiệp hoặc bán buôn. Điều này có thể lý giải vì sao các nước châu âu và Úc là nước tạo doanh thu nhiều nhất do có sự tương đồng trong xu hướng văn hóa đã góp phần thúc đẩy doanh thu.

Ta xác định được tỷ lệ khách hàng chỉ mua 1 lần chiếm 24,57% và tỷ lệ khách hàng đóng góp 80% doanh thu chiếm 22.74%.

Hình 3‑15. Biểu đồ tỷ lệ khách hàng mua 1 lần và nhiều lần

Ta nhận thấy rằng 24.75% khách hàng chỉ mua một lần, đây là nhóm khách hàng tiềm năng có thể được tiếp cận lại thông qua các chiến lược tiếp thị hoặc gợi ý sản phẩm để tăng tần suất mua sắm của họ.

Đồng thời, 80% doanh thu được đóng góp bởi chỉ 22.74% khách hàng, kết quả này gần với nguyên tắc Pareto(80/20), cho thấy 22.74% khách hàng có vai trò quan trọng và là nguồn đóng góp chính cho doanh thu của công ty.

Dựa trên những kết quả này, nhóm quyết định xây dựng hệ thống gợi ý cho người tiêu dùng với mô hình NCF, không chỉ nhằm tăng sự hài lòng của khách hàng hiện tại mà còn để khuyến khích khách hàng mua sắm nhiều hơn. Điều này sẽ giúp tối ưu hóa doanh thu và nâng cao trải nghiệm khách hàng.

## Huấn luyện mô hình học sâu

* Xử lý dữ liệu:

Nhóm sẽ chỉ dùng 4 đặc trưng trong 8 đặc trưng bao gồm ‘Customer ID’, StockCode, ‘Quantity’, ‘InvoiceDate’ để xây dựng mô hình. Đồng thời chuyển kiểu dữ liệu cột InvoiceDate sang dạng datetime và 3 cột còn lại sang dạng số nguyên.

Sau đó nhóm loại bỏ các giá trị NaN trong cột Customer ID, những giá trị âm trong cột Quantity và những mã hóa đơn hợp lệ sẽ được giữ lại.

Lọc cộng tác dựa trên phản hồi ngầm tập trung vào việc dự đoán sở thích của người dùng dựa trên lịch sử tương tác của họ với các sản phẩm. Trong trường hợp này, phản hồi ngầm là các dấu hiệu như số lượng mua hàng (cột Quantity), thay vì xếp hạng rõ ràng từ khách hàng.

Dữ liệu trong cột Quantity được chuyển thành giá trị nhị phân:

1: Khách hàng đã mua hàng (Quantity > 0).

0: Khách hàng chưa mua hàng (Quantity = 0).

Bảng 3‑2. Bảng thống kê độ thưa của bộ dữ liệu

|  |  |  |  |
| --- | --- | --- | --- |
| **Số giao dịch** | **Số sản phẩm** | **Số khách hàng** | **Độ thưa** |
| **802663** | **3862** | **5850** | **96,44%** |

Bộ dữ liệu rất thưa và mức thưa thớt của ma trận này là 96,44%, trong đó hơn 20% khách hàng chỉ có 1 giao dịch, điều này gây khó khăn khi đánh giá mô hình lọc cộng tác nên nhóm chỉ giữ lại khách hàng có ít nhất 10 giao dịch.

* Chia tập dữ liệu

Mô hình này dùng để dự đoán hành vi mua hàng trong tương lai dựa trên dữ liệu lịch sử của khách hàng. Nhóm sử dụng phương pháp LOOV đã đề cập ở chương 2 nên bộ dữ liệu được chia thành hai tập, tập train và tập test dựa trên thời gian giao dịch của khách hàng.

+ Tập train bao gồm tất cả giao dịch ngoại trừ giao dịch gần nhất của mỗi khách hàng

+ Tập test chỉ bao gồm giao dịch gần nhất của mỗi khách hàng

* Tạo mẫu âm tính

Với mỗi khách hàng, các sản phẩm mà khách hàng chưa tương tác là các sản phẩm tiềm năng. Và trong bộ dữ liệu này mới chỉ đang có các giao dịch mà khách hàng đã thực hiện tức là chỉ có mẫu dương tính. Mẫu âm tính đại diện cho các mặt hàng mà khách hàng chưa tương tác và nhóm lấy mẫu âm tính các sản phẩm ngẫu nhiên cho từng khách hàng để tránh thiên vị. Tỷ lệ giữa mẫu dương tính và mẫu âm tính là 1:4.

* Mô hình NeuMF:

Cấu trúc mô hình được triển khai như sau:

**- GMF:**

+ Tạo các embedding MF riêng biệt cho khách hàng và sản phẩm.

+ Sau đó nhân từng phần tử giữa embedding của khách hàng và sản phẩm để tạo vector đặc trưng kết hợp.

**- MLP:**

+ Tạo các embedding MLP riêng biệt cho khách hàng và sản phẩm.

+ Nối embedding của khách hàng và sản phẩm thành một vector trước khi đưa vào các lớp MLP.

+ Kiến trúc MLP:

Bao gồm 3 lớp fully connected.

Số nơron của mỗi lớp giảm dần theo hình tháp, với số nơron lớp sau bằng một nửa số nơron của lớp trước.

Các lớp sử dụng hàm kích hoạt ReLU, trừ lớp cuối cùng.

Kết hợp thêm lớp Dropout để giảm khả năng overfitting

- Sau đó nối đầu ra của GMF với đầu ra lớp cuối cùng của MLP và đưa qua lớp fully connected cuối cùng rồi dùng hàm kích hoạt Sigmoid để dự đoán xác suất khách hàng có quan tâm đến sản phẩm.

- Mô hình sử dụng hàm mất mát Binary Cross Entropy để tối ưu hóa dự đoán.

- Sử dụng Adam để tối ưu trọng số và bias.

## Đánh giá và so sánh mô hình

Nhằm giảm thiểu thời gian tính toán mà vẫn đảm bảo đánh giá chính xác hiệu suất mô hình, nhóm đã áp dụng chiến lược xếp hạng 100 sản phẩm gần nhất đối với mỗi khách hàng gồm 99 sản phẩm ngẫu nhiên chưa tương tác và 1 sản phẩm đã tương tác. Chọn top 10 sản phẩm từ danh sách 100 sản phẩm sau khi chạy mô hình.

Nhóm sử dụng 2 chỉ số HR@10 và NDCG@10 để đánh giá hiệu suất mô hình với 4 bộ tham số chính bao gồm:

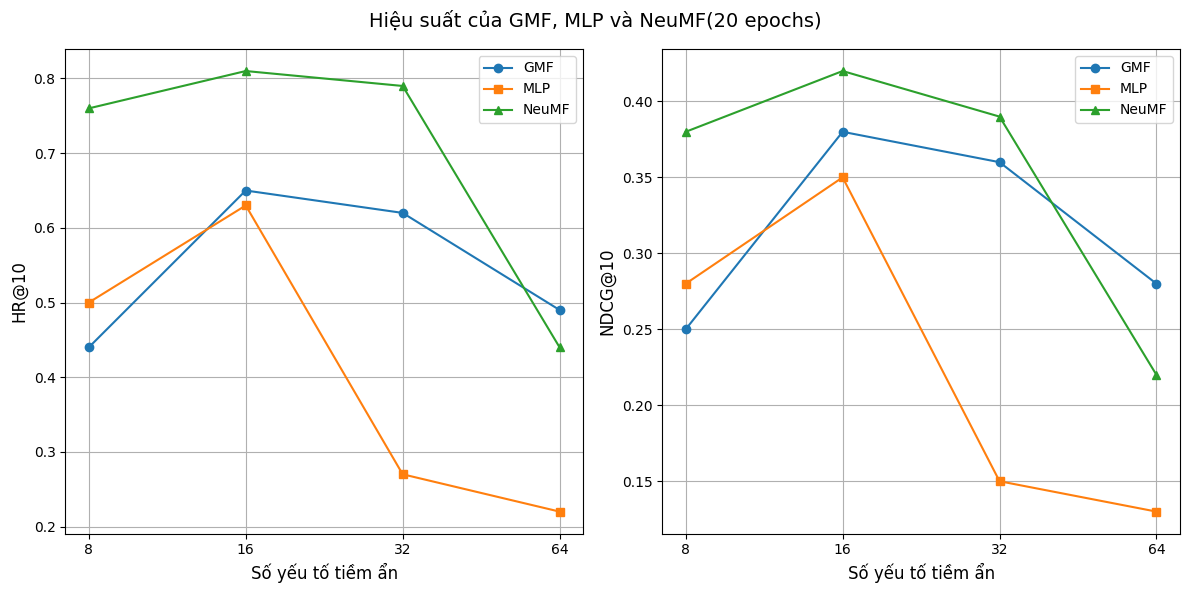
* Số yếu tố tiềm ẩn: 8, 16, 32, 64
* Tốc độ học: 0.0001, 0.0005, 0.001, 0.005
* Kích thước batch: 64, 128, 256, 512
* Lớp ẩn dựa trên số yếu tố tiềm ẩn:

+ Với yếu tố tiềm ẩn = 8: [16, 8, 4]

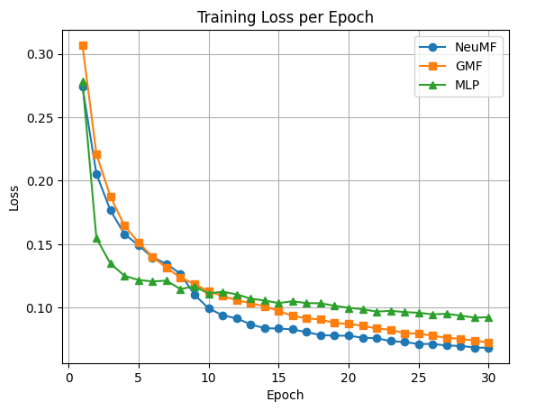
+ Với yếu tố tiềm ẩn = 16: [32, 16, 8]

+ Với yếu tố tiềm ẩn = 32: [64, 32, 16]

+ Với yếu tố tiềm ẩn = 64: [128, 64, 32]

Mô hình NeuMF có hiệu suất cao nhất trong cả hai chỉ số HR@10 và NDCG@10, đặc biệt ở mức số yếu tố tiềm ẩn là 16 đạt giá trị cao nhất với HR@10(0.81) và NDCG@10(0.42). Tuy nhiên khi số yếu tố tiềm ẩn tăng lên(32 và 64) thì hiệu suất của cả 3 mô hình đều giảm mạnh. Điều này cho thấy mô hình có xu hướng bị overfitting khi tăng số yếu tố tiềm ẩn lên quá cao.

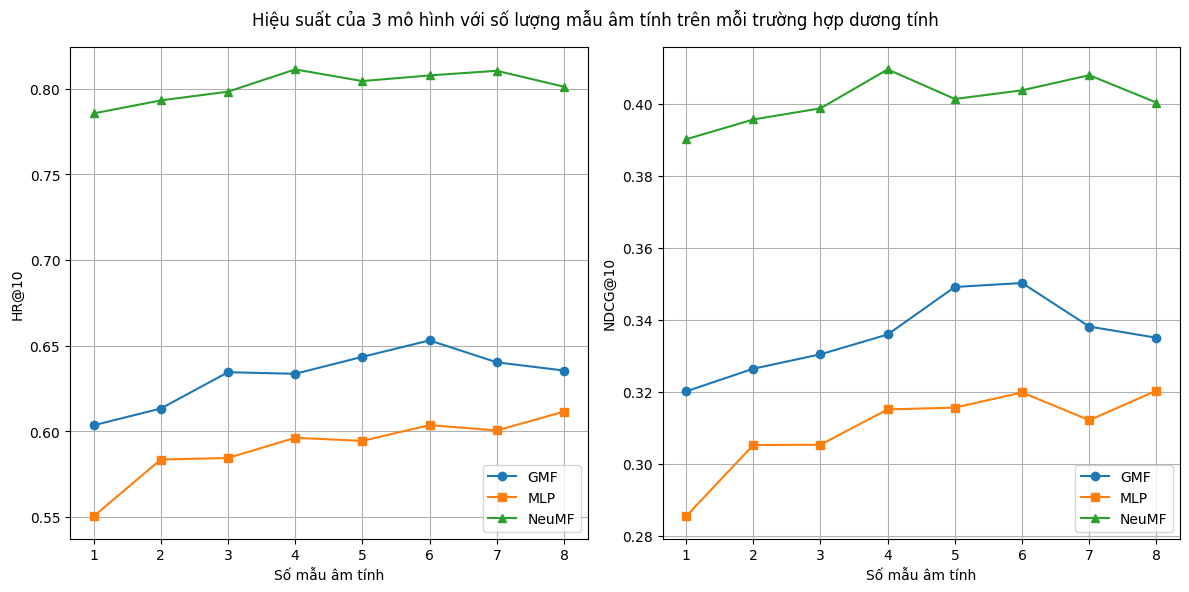
Hình 3‑16. Hiệu suất của 3 mô hình

Xu hướng của giá trị mất mát trong quá trình huấn luyện trong 10 lần huấn luyện đầu tiên đều giảm rõ rệt, sau đó trở về trạng thái ổn định. GMF có tốc độ giảm nhanh từ 1 đến 5 epoch và chậm dần sau 10 epoch. MLP có tốc độ giảm nhanh nhất trong giai đoạn 1 đến 3 epoch, tuy nhiên sau 5 epoch thì giảm chậm lại và có mức hội tụ cao nhất và cũng là mô hình có hiệu suất thấp nhất. NeuMF có giá trị mất mát thấp nhất và đạt hiệu suất cao nhất trong 3 mô hình.

Hình 3‑17. Biểu đồ giá trị mất mát của 3 mô hình trong huấn luyện

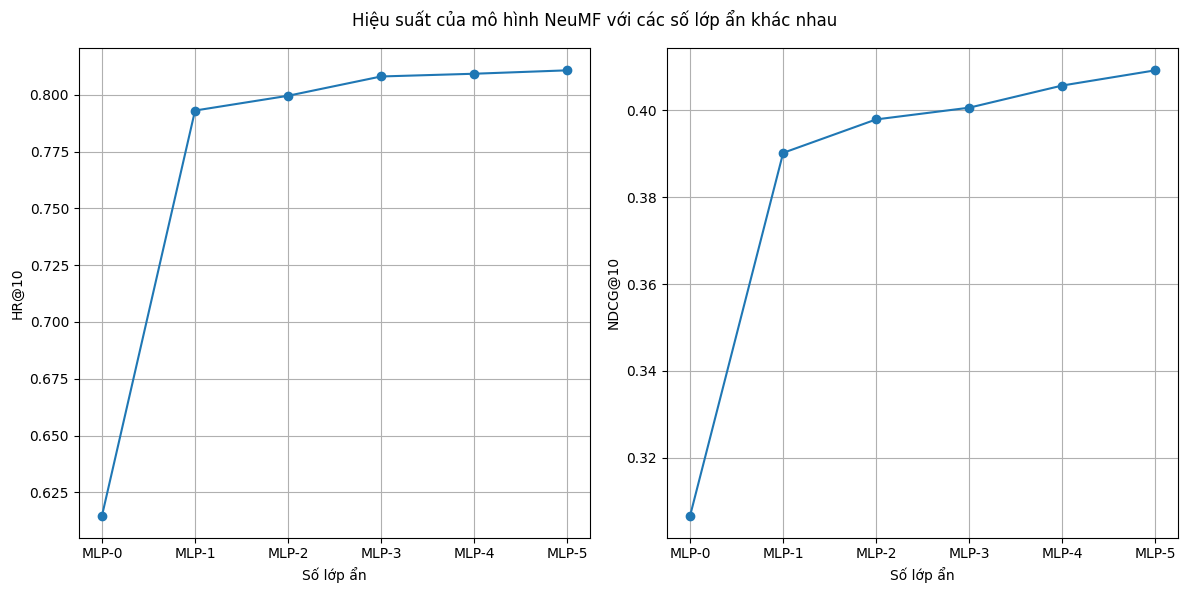
Tóm lại, NeuMF là mô hình tốt nhất về cả hiệu suất và khả năng tối ưu hóa loss. Điều này cho thấy NeuMF tận dụng tốt ưu điểm khi kết hợp 2 mô hình GMF và MLP và tính hợp lý, đúng đắn và hiệu quả khi dùng BCE trong việc học qua dữ liệu ngầm.

Hình 3‑18. Hiệu suất 3 mô hình với số lượng mẫu âm tính trên mỗi trường hợp dương tính



Ta thấy rằng việc lấy 1 mẫu dương tính và 1 mẫu âm tính cho kết quả thấp nhất. Tỷ lệ lấy mẫu tối ưu là khoảng từ 3 đến 6. Khi tỷ lệ lấy mẫu lớn hơn 6, hiệu suất của mô hình bắt đầu giảm. Có thể kết luận rằng việc chọn tỷ lệ lấy mẫu quá lớn có thể ảnh hưởng xấu đến hiệu suất của mô hình

Hình 3‑19. Hiệu suất của NeuMF với các lớp ẩn khác nhau trong MLP

Để đánh giá hiệu suất mô hình khi cài đặt các lớp ẩn khác nhau trong mô hình MLP, ta thấy rằng khi mô hình NeuMF không có lớp MLP nào(mô hình tuyến tính) thì mô hình đạt hiệu suất thấp nhất. Điều này chứng tỏ tính phi tuyến của mô hình MLP mang lại là nguyên nhân giúp cải thiện hiệu suất cho mô hình NeuMF. Bên cạnh đó khi tăng dần số lớp ẩn, mô hình cũng cải thiện hiệu suất lên một chút.

# KẾT LUẬN

* Các kết quả đã đạt được:

Với hướng tiếp cận nghiên cứu về việc kết hợp phân rã ma trận với mạng nơron vào lĩnh vực hệ thống gợi ý, đề tài “Xây dựng hệ thống gợi ý cho người người tiêu dùng dựa trên mô hình học sâu” đã đạt được những kết quả quan trọng sau:

+ Kiến thức tổng quát:

Đề tài đã cung cấp một cái nhìn tổng quan về hệ thống gợi ý và các phương pháp tiếp cận cơ bản để xây dựng hệ thống gợi ý.

+ Hệ thống lý thuyết:

Đã thực hiện việc hệ thống hóa các cơ sở lý thuyết về hệ thống gợi ý, tập trung phân tích và tổng hợp các nghiên cứu liên quan để xây dựng hệ thống gợi ý dựa trên lọc cộng tác.

+ Xây dựng mô hình:

Một mô hình kết hợp phương pháp phân rã ma trận với mạng nơron và lọc cộng tác đã được xây dựng thành công. Mô hình được đánh giá dựa trên các chỉ số quan trọng HR và NDCG đã mang lại kết quả gần với thực tế và đáp ứng tốt yêu cầu bài toán.

* Định hướng mở rộng và phát triển đề tài trong tương lai:

Để cải thiện hiệu suất của mô hình chúng ta có thể mở rộng thêm bộ dữ liệu, tối ưu hóa các tham số đào tạo. Kết hợp thêm các phương pháp gợi ý dựa trên nội dung để tạo ra mô hình lai, đồng thời sử dụng thêm các mô hình phân cụm để phân cụm khách hàng nhằm nâng cao độ chính xác và tính cá nhân hóa cho khách hàng.