



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KHOA HỌC & KỸ THUẬT THÔNG TIN

Hotels Recommendation System

SVTH: Team 10

- Phạm Đức Thế
- Trần Thành Luân

GVHD:

- Ths. Nguyễn Văn Kiệt
- CN. Huỳnh Văn Tín

NỘI DUNG

- ☒ Giới thiệu
- ☐ Bộ dữ liệu
- ☐ Phương pháp
- ☐ Đánh giá
- ☐ Demo
- ☐ Kết luận

GIỚI THIỆU

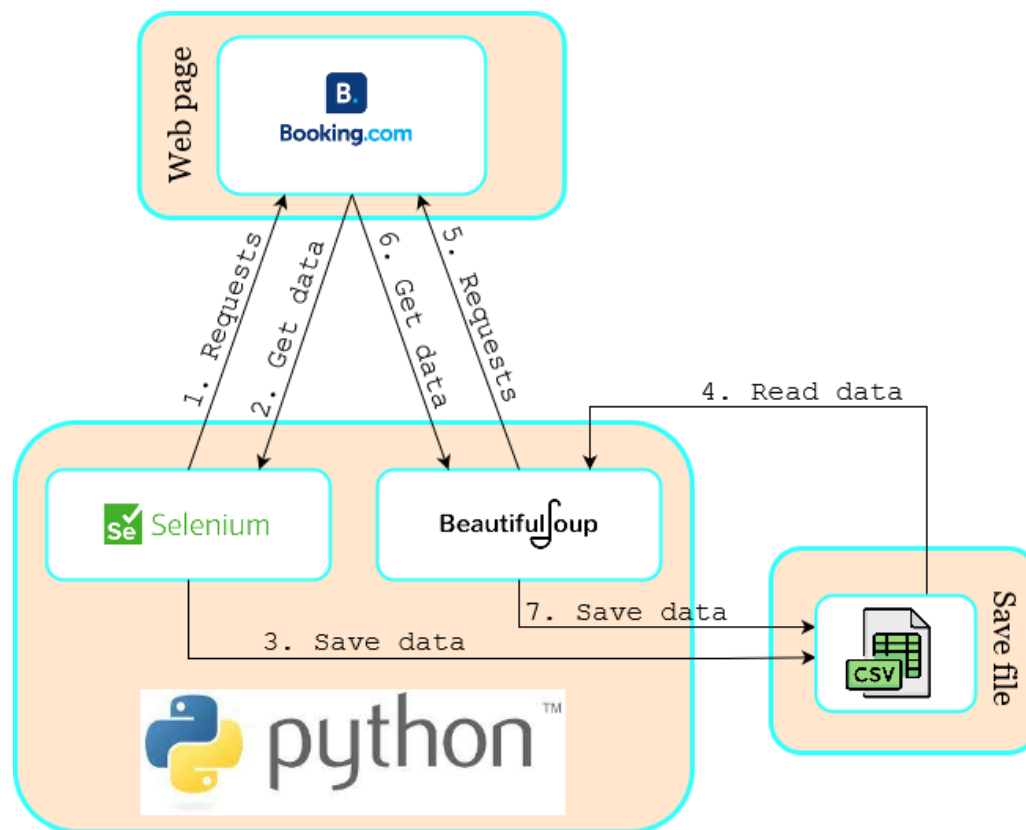
- Hệ thống Khuyến nghị (Recommendation System) là một mảng khá rộng của machine learning, mục đích chính của các recommendation system là dự đoán mức độ quan tâm của một người dùng tới một sản phẩm nào đó.
- Với sự phát triển mạnh mẽ của du lịch và internet, nhu cầu đặt phòng khách sạn trực tuyến càng ngày càng lớn, việc tìm kiếm thông tin khách sạn ngày càng trở nên khó khăn và mất khá nhiều thời gian.
- Xây dựng hệ khuyến nghị khách sạn, giúp cho việc tìm kiếm thông tin khách sạn trở nên dễ dàng và tối ưu hóa trải nghiệm người dùng trên các trang web đặt chỗ khách sạn du lịch trực tuyến.

NỘI DUNG

- ☐ Giới thiệu
- ☒ Bộ dữ liệu
- ☐ Phương pháp
- ☐ Đánh giá
- ☐ Demo
- ☐ Kết luận

BỘ DỮ LIỆU

□ Crawl data



BỘ DỮ LIỆU

□ Crawl data

Index	Thuộc tính	Ý nghĩa
0	URL Hotel	URL của khách sạn.
1	Location	Địa điểm của khách sạn (Tên tỉnh/TP).
2	HotelID	ID của khách sạn.
3	Name Hotel	Tên của khách sạn.
4	Descriptions	Mô tả, giới thiệu của khách sạn.
5	Address	Địa chỉ của khách sạn.
6	UserID	ID của user.
7	User	Tên của user (người đánh giá).
8	Rating	Rating của user.

Bảng 1: Thông tin chi tiết của các thuộc tính.

BỘ DỮ LIỆU

□ Crawl data

	URL Hotel	Location	HotelID	Name Hotel	Descriptions	Address	UserID	User	Rating
0	https://www.booking.com/hotel/vn/dalat-wind.vi...	Đà Lạt	4064	Dalat Wind Deluxe Hotel	Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hươ...	Lot R2 03-04. Golf Valley, Ward 2, Đà Lạt, Viê...	1187	Thảo	6
1	https://www.booking.com/hotel/vn/dalat-wind.vi...	Đà Lạt	4064	Dalat Wind Deluxe Hotel	Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hươ...	Lot R2 03-04. Golf Valley, Ward 2, Đà Lạt, Viê...	1284	Tran	5
2	https://www.booking.com/hotel/vn/dalat-wind.vi...	Đà Lạt	4064	Dalat Wind Deluxe Hotel	Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hươ...	Lot R2 03-04. Golf Valley, Ward 2, Đà Lạt, Viê...	5866	Tho	6
3	https://www.booking.com/hotel/vn/dalat-wind.vi...	Đà Lạt	4064	Dalat Wind Deluxe Hotel	Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hươ...	Lot R2 03-04. Golf Valley, Ward 2, Đà Lạt, Viê...	3033	Tuan	9
4	https://www.booking.com/hotel/vn/dalat-wind.vi...	Đà Lạt	4064	Dalat Wind Deluxe Hotel	Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hươ...	Lot R2 03-04. Golf Valley, Ward 2, Đà Lạt, Viê...	1406	Phan	7
...
38796	https://www.booking.com/hotel/vn/littleboss.vi...	Huế	135	LittleBoss	Căn hộ có TV màn hình phẳng. Căn hộ nằm trong...	The Manor Crown Đường Tổ Hữu, TP Huế, Huế, Vi...	2190	Linh	6
38797	https://www.booking.com/hotel/vn/littleboss.vi...	Huế	135	LittleBoss	Căn hộ có TV màn hình phẳng. Căn hộ nằm trong...	The Manor Crown Đường Tổ Hữu, TP Huế, Huế, Vi...	1530	Anh	9
38798	https://www.booking.com/hotel/vn/littleboss.vi...	Huế	135	LittleBoss	Căn hộ có TV màn hình phẳng. Căn hộ nằm trong...	The Manor Crown Đường Tổ Hữu, TP Huế, Huế, Vi...	1308	Phong	7
38799	https://www.booking.com/hotel/vn/littleboss.vi...	Huế	135	LittleBoss	Căn hộ có TV màn hình phẳng. Căn hộ nằm trong...	The Manor Crown Đường Tổ Hữu, TP Huế, Huế, Vi...	6009	Hạnh	6
38800	https://www.booking.com/hotel/vn/littleboss.vi...	Huế	135	LittleBoss	Căn hộ có TV màn hình phẳng. Căn hộ nằm trong...	The Manor Crown Đường Tổ Hữu, TP Huế, Huế, Vi...	5564	Hiền	8

38801 rows × 9 columns

BỘ DỮ LIỆU

□ Thống kê mô tả

Thuộc tính	Thông tin
Location	Gồm 10 địa điểm: Đà Lạt, Hà Nội, Đà Nẵng,...
HotelID	Gồm 4506 hotels.
UserID	Gồm 6471 users.
Rating	Được đánh giá từ 1-10.

BỘ DỮ LIỆU

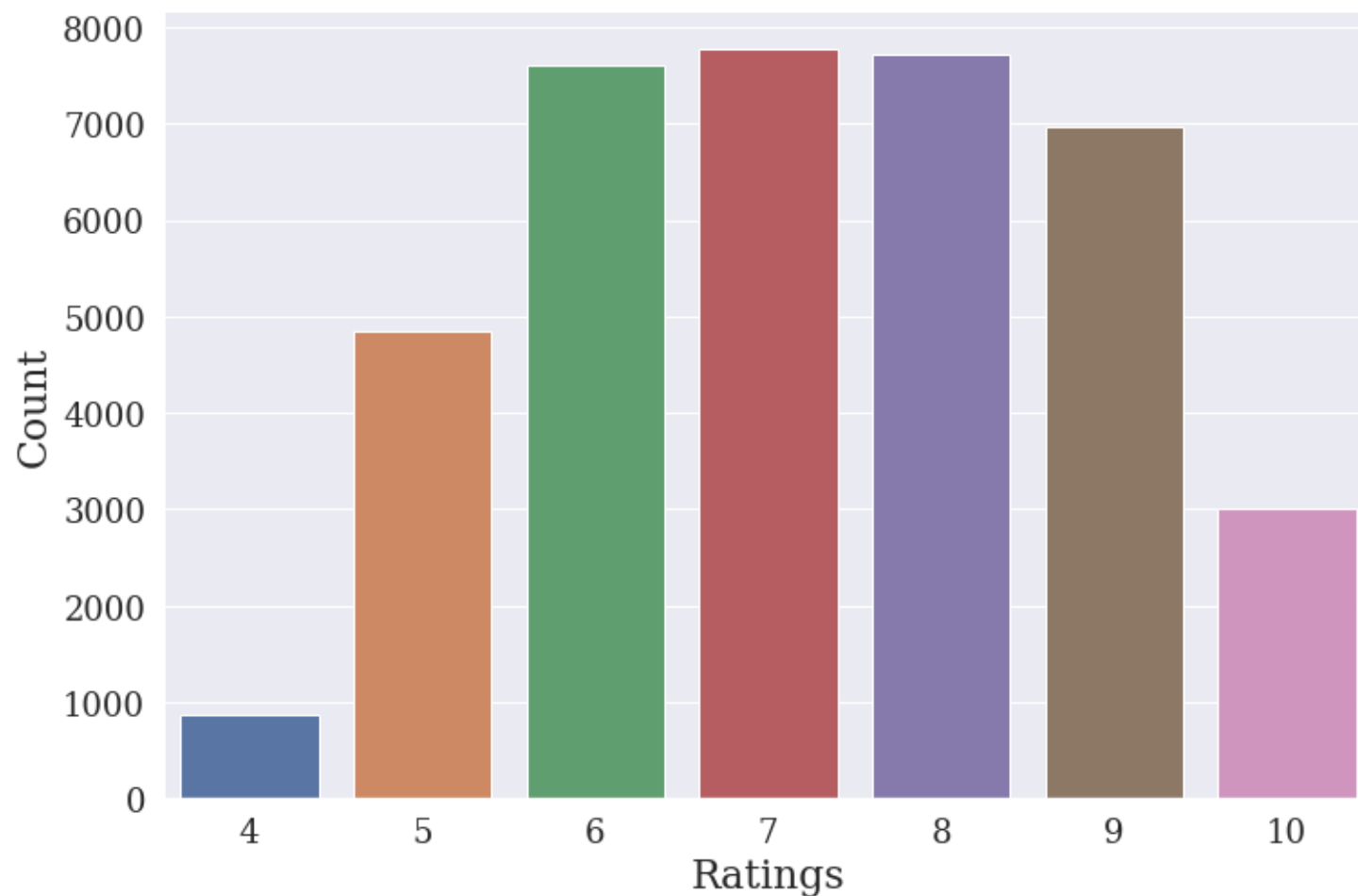
□ Thống kê mô tả

Ví dụ về Descriptions của một khách sạn:

Tọa lạc tại thành phố Đà Lạt, cách Hồ Xuân Hương 500 m, Dalat Wind Deluxe Hotel là khách sạn 2 sao có lễ tân 24 giờ, sảnh khách chung, Wi-Fi và chỗ đỗ xe riêng miễn phí. Tại khách sạn, tất cả các phòng đều có bàn làm việc, TV màn hình phẳng, ấm đun nước và phòng tắm riêng với chậu rửa vệ sinh (bidet). Một số phòng còn có ban công. Khách sạn phục vụ bữa sáng gọi món hàng ngày. Du khách có thể dùng bữa tại nhà hàng trong khuôn viên, nơi chuyên phục vụ các món nướng/BBQ. Du khách cũng có thể thư giãn trên sân hiên tắm nắng. Các điểm tham quan nổi tiếng gần Dalat Wind Deluxe Hotel bao gồm Vườn hoa Đà Lạt, Quảng trường Lâm Viên và Công viên Yersin Đà Lạt. Sân bay gần nhất là sân bay Liên Khương, cách chỗ nghỉ 23 km.

BỘ DỮ LIỆU

□ Thống kê mô tả



NỘI DUNG

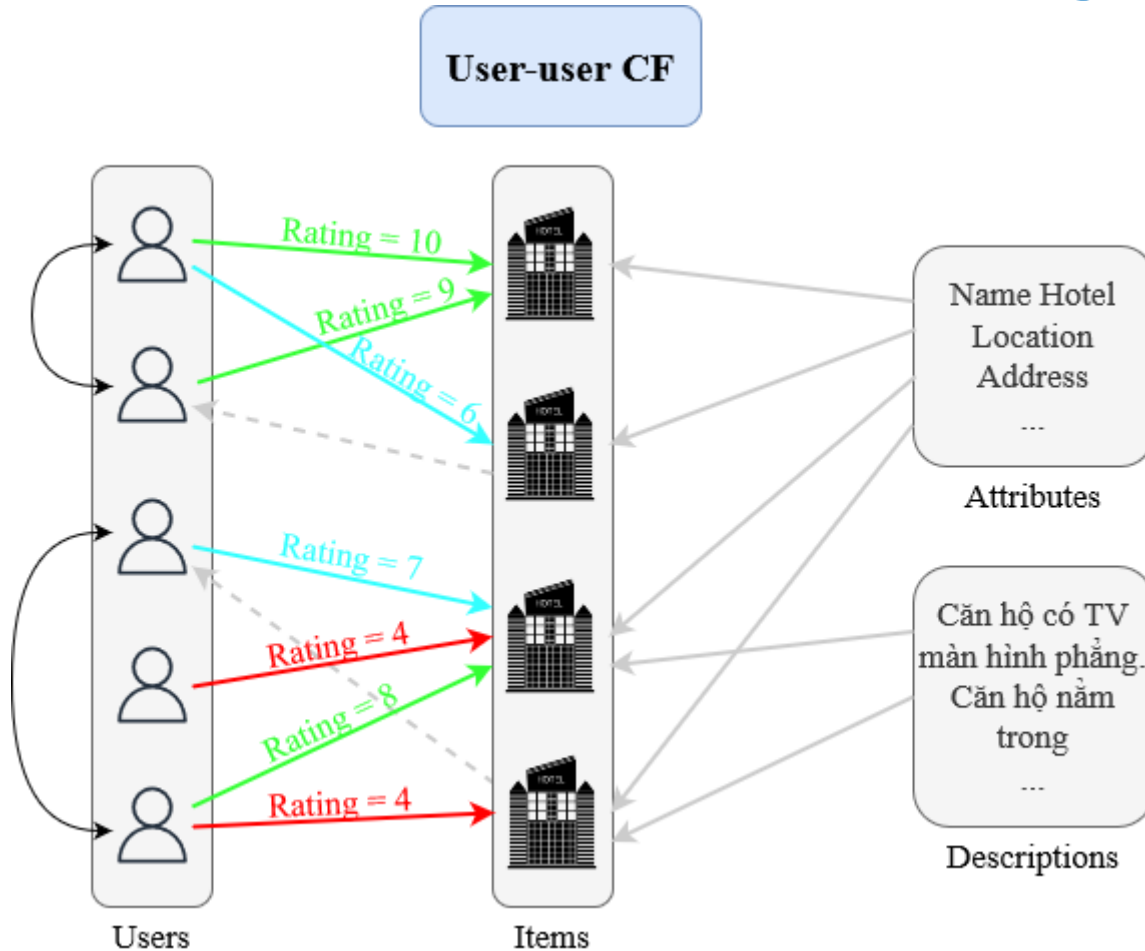
- ☐ Giới thiệu
- ☐ Bộ dữ liệu
- ☒ Phương pháp
- ☐ Đánh giá
- ☐ Demo
- ☐ Kết luận

PHƯƠNG PHÁP

- ❑ Collaborative Filtering.
- ❑ Content-based Filtering.

PHƯƠNG PHÁP

❑ User-user collaborative filtering.



Input: Ma trận đánh giá user-rating (rating matrix)

- Dựa vào độ tương đồng giữa các user u .
- Sử dụng độ tương đồng cosine và pearson.

Output:

- Top-N item được khuyến nghị.
(Dự đoán mức độ quan tâm của user u cho item i .)

PHƯƠNG PHÁP

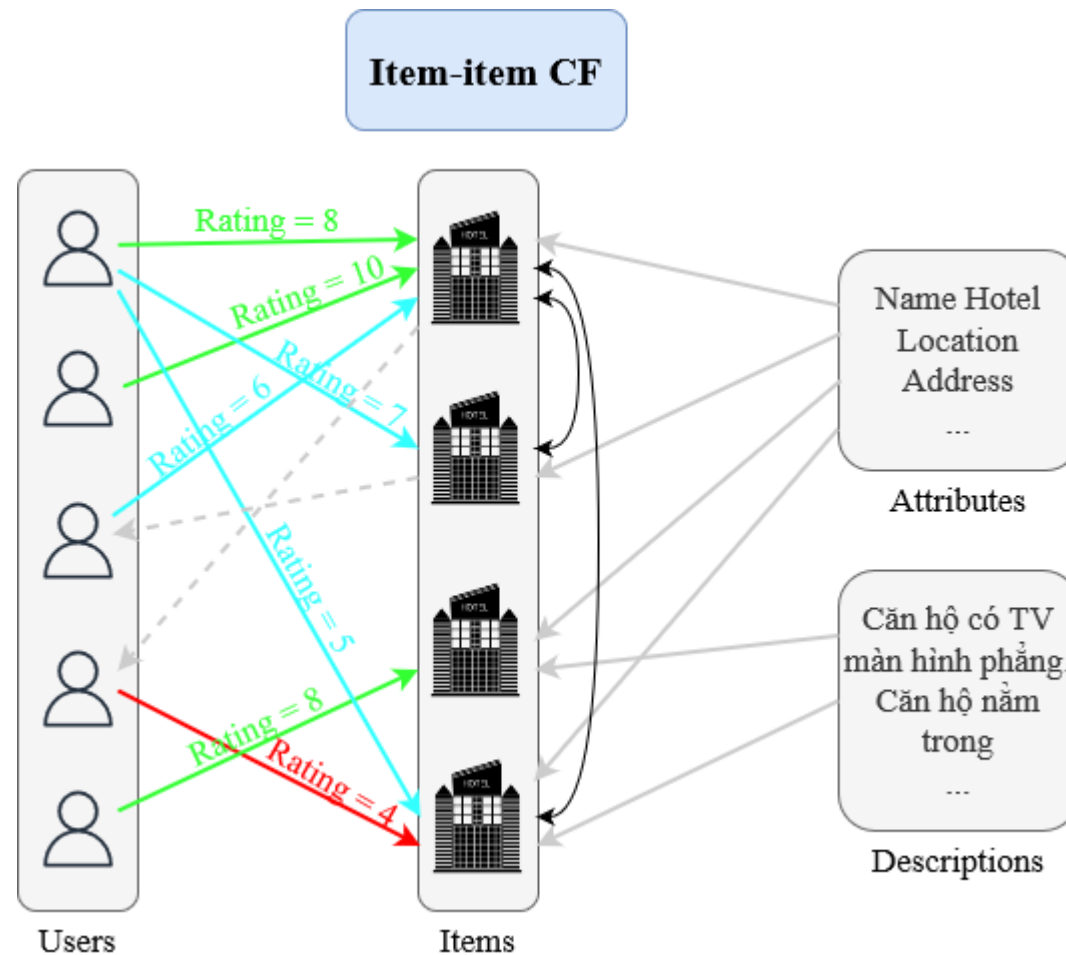
❑ Item-item collaborative filtering.

Input: Ma trận đánh giá item-rating (rating matrix)

- Dựa vào độ tương đồng giữa các item i .
- Sử dụng độ tương đồng cosine và pearson

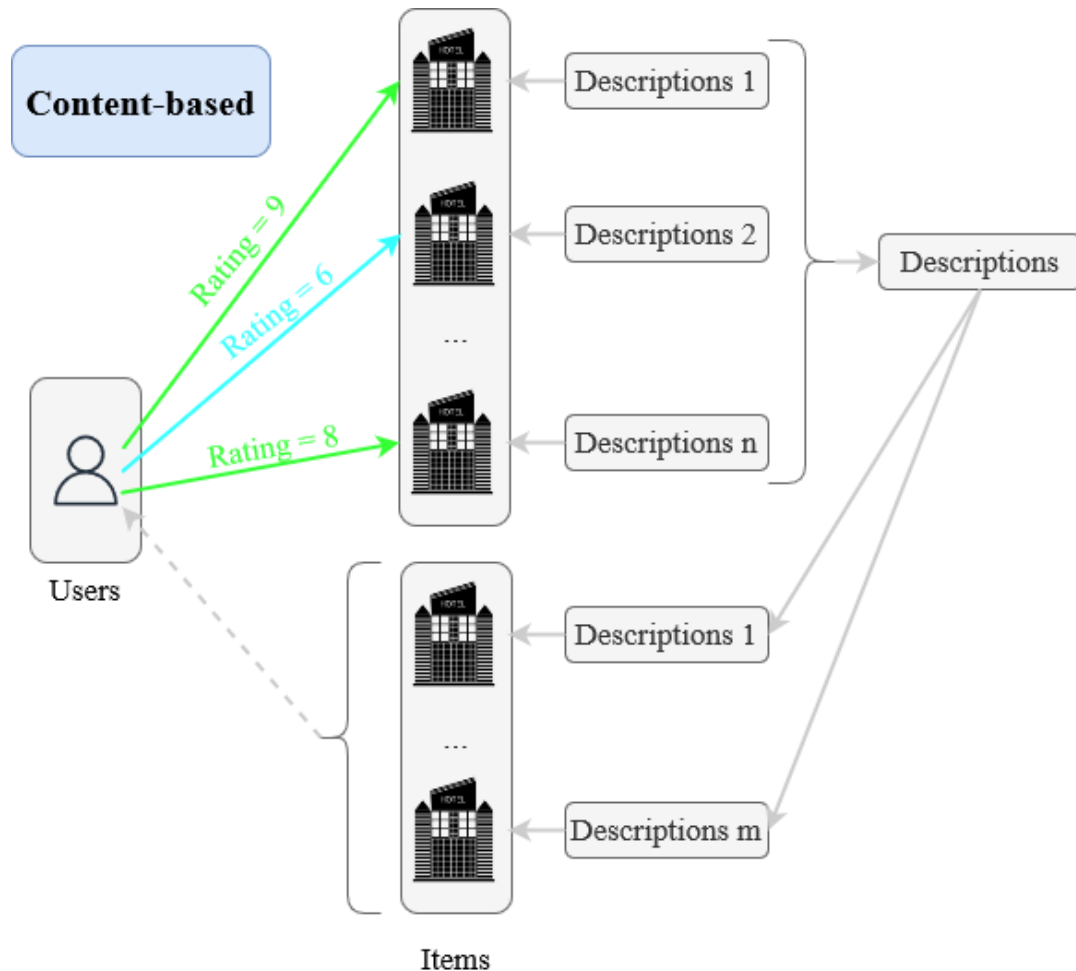
Output:

- Top-N những item được khuyến nghị.
(Dự đoán mức độ phù hợp của item i với user u).



PHƯƠNG PHÁP

□ Content-based filtering.



Input: Mô tả của tất cả các khách sạn.

- Merge tất cả các mô tả của khách sạn mà người dùng đã đi thành bộ mô tả cho người dùng.
- TF-IDF
- Sử dụng độ tương đồng cosine.
- Tìm ra điểm tương đồng giữa bộ mô tả của người dùng và tất cả các hotel hiện có.

Output:

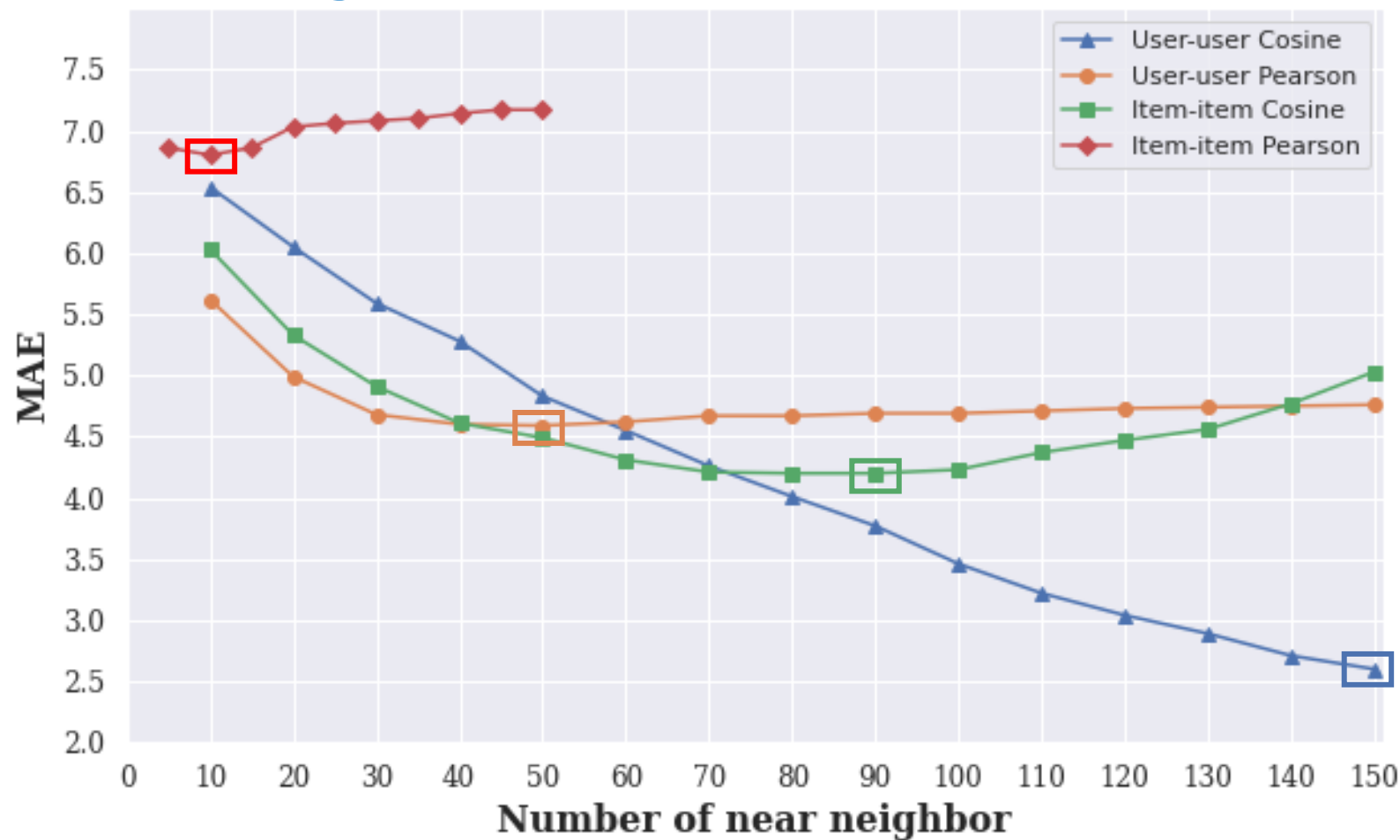
- Top-N item được khuyến nghị.

NỘI DUNG

- ☐ Giới thiệu
- ☐ Bộ dữ liệu
- ☐ Phương pháp
- ☒ **Đánh giá**
- ☐ Demo
- ☐ Kết luận

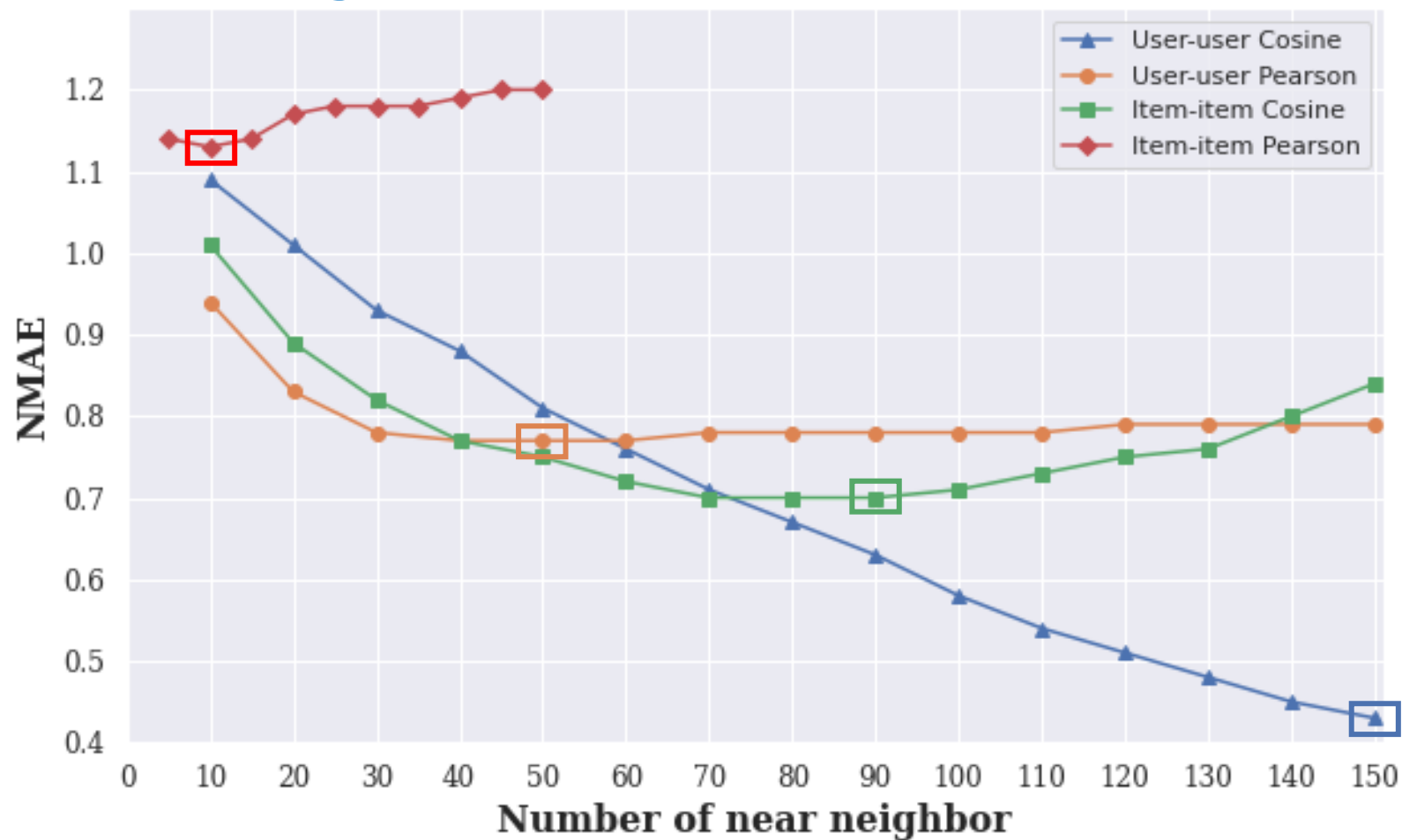
ĐÁNH GIÁ

Collaborative filtering.



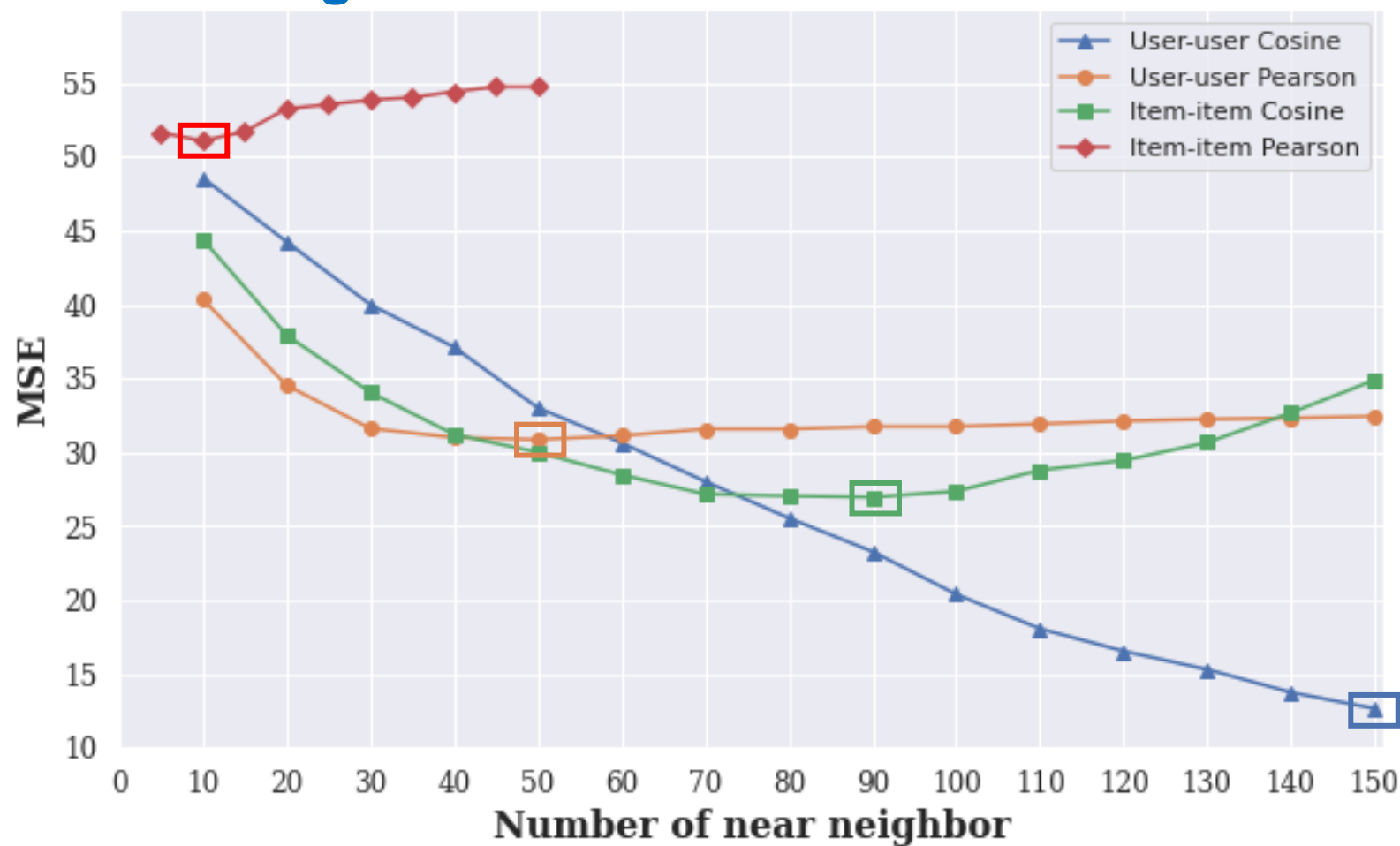
ĐÁNH GIÁ

Collaborative filtering.



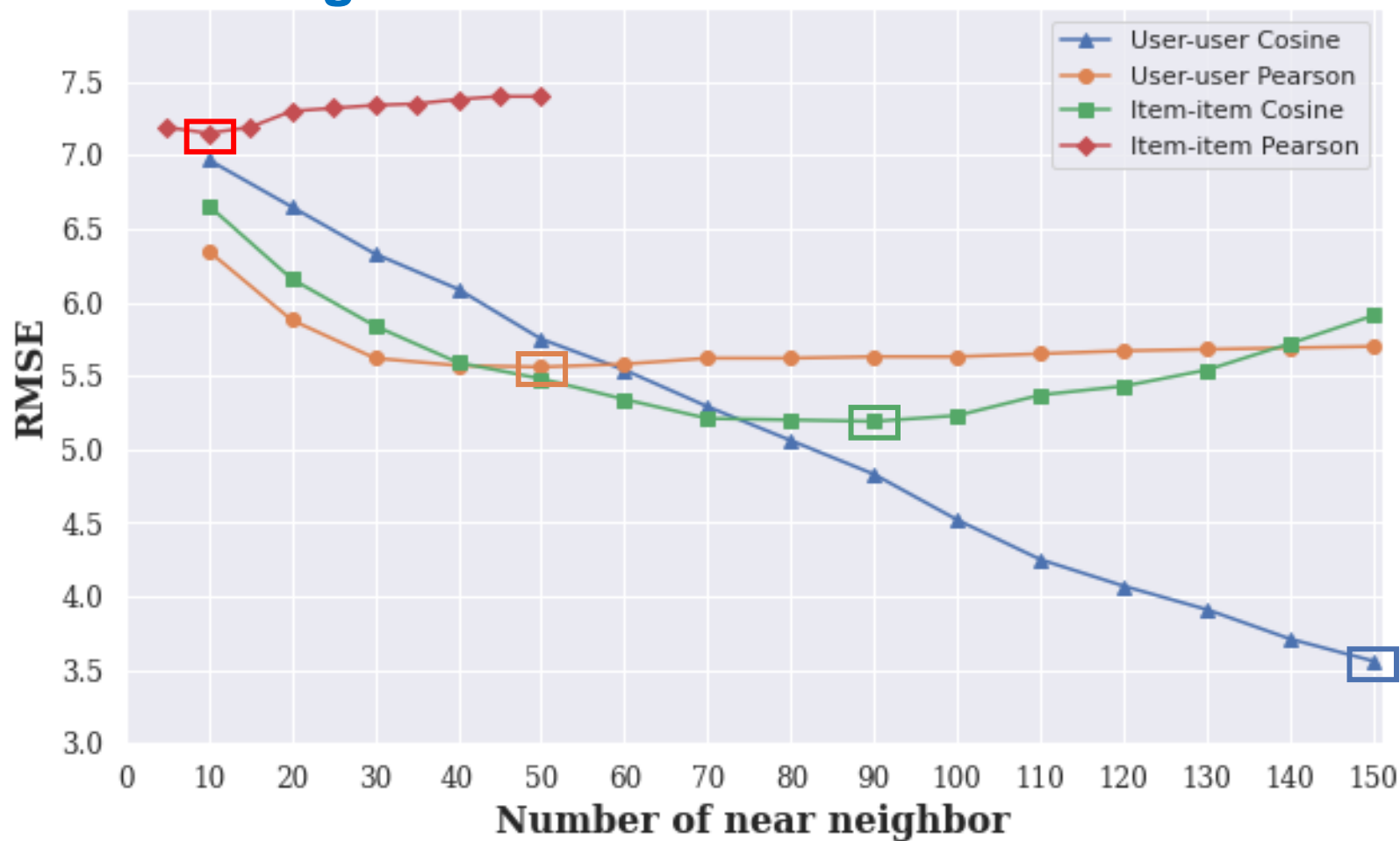
ĐÁNH GIÁ

Collaborative filtering.



ĐÁNH GIÁ

Collaborative filtering.



ĐÁNH GIÁ

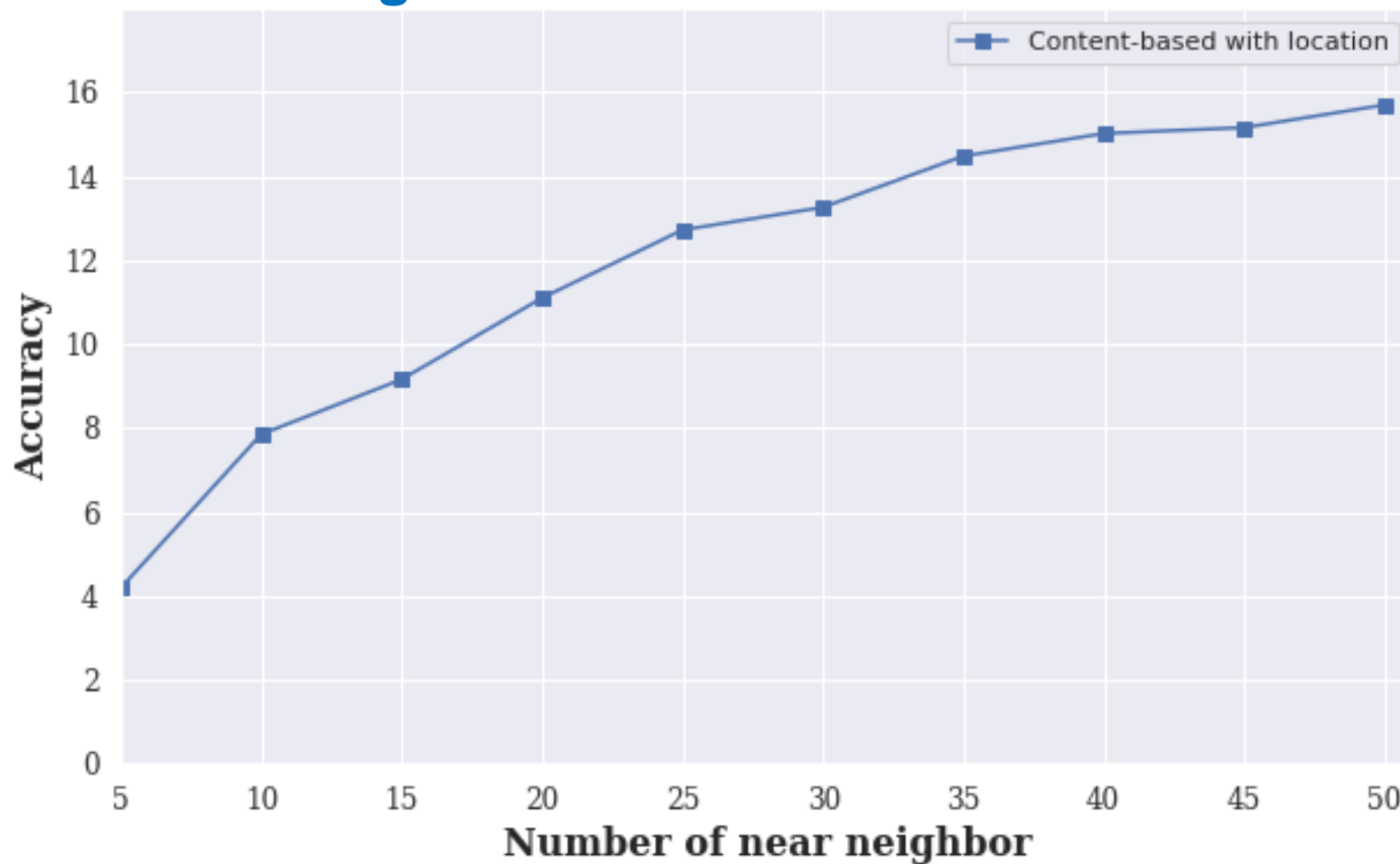
□ Collaborative filtering.

		Độ đo			
		MSE	RMSE	MAE	NMAE
Mô hình	user-user cosine	12.66	3.56	2.60	0.43
	user-user pearson	30.87	5.56	4.59	0.77
	item-item cosine	26.97	5.19	4.20	0.70
	item-item pearson	51.09	7.15	6.80	1.13

Bảng 2: Kết quả tốt nhất của các mô hình theo các độ đo.

ĐÁNH GIÁ

□ Content-based filtering.



NỘI DUNG

- ☐ Giới thiệu
- ☐ Bộ dữ liệu
- ☐ Phương pháp
- ☐ Đánh giá
- ☒ **Demo**
- ☐ Kết luận

Demo

https://youtu.be/GsxB_sPU-DE

NỘI DUNG

- ☐ Giới thiệu
- ☐ Bộ dữ liệu
- ☐ Phương pháp
- ☐ Đánh giá
- ☐ Demo
- ☒ **Kết luận**

KẾT LUẬN

- Chúng tôi đã thu thập, xây dựng và trình bày bộ dữ liệu *Booking Hotels Dataset*, một bộ dữ liệu mới cho bài toán khuyến nghị đặt phòng khách sạn du lịch trực tuyến.
- Với phương pháp collaborative filtering chúng tôi đã cài đặt thành công bốn mô hình memory-based gồm: user-user cosine, user-user pearson, item-item cosine và item-item pearson; với phương pháp content-based filtering chúng tôi cũng đã cài đặt thành công mô hình content-based.
- Kết quả tốt nhất mà chúng tôi đạt được là 12.66 MSE, 3.56 RMSE, 2.60 MAE và 0.34 NMAE sử dụng mô hình user-user cosine.

KẾT LUẬN

□ Hướng phát triển.

- **Bộ dữ liệu:** Thu thập thêm dữ liệu từ các trang web đặt phòng khách sạn du lịch trực tuyến, xử lý hiện tượng thừa thớt trong các rating matrix.
- **Mô hình:** Áp dụng các phương pháp, kỹ thuật khuyến nghị khác như: Collaborative Filtering dùng model-based, Knowledge-Based Recommender Systems, Demographic Recommender Systems, Hybrid and Ensemble-Based Recommender Systems . . . để cải thiện kết quả dự đoán tốt hơn nữa.

THANKS FOR YOUR ATTENTION!

QUESTION & ANSWER

