

# **H&M Fashion Recommendation system for Existing customers and New Customers**

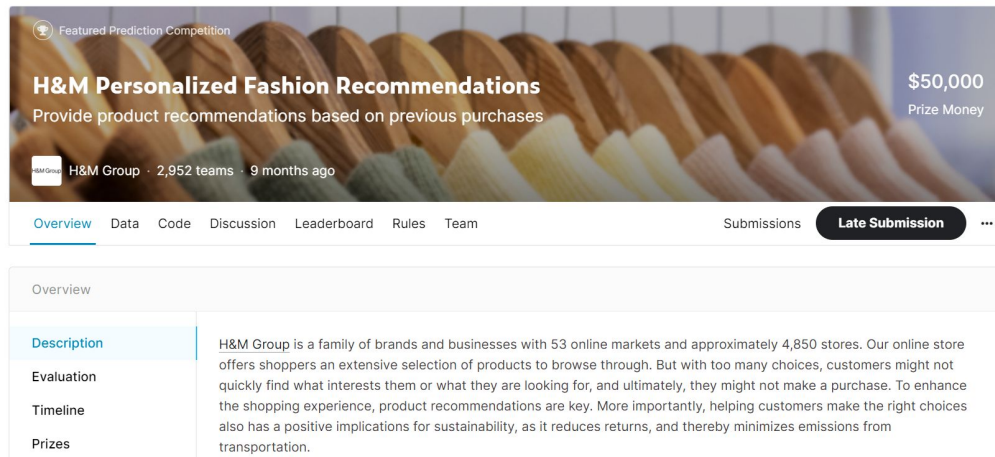
**GVHD:** TS. Đỗ Trọng Hợp

**Sinh viên thực hiện:**

1. Nguyễn Lê Thanh - 19522238
2. Đinh Thị Diễm Sương - 19522145

# Giới thiệu về đề tài

- Đây là một cuộc thi trên Kaggle, nhiệm vụ là dự đoán mười hai sản phẩm `article_ids` mà mỗi khách hàng có thể mua trong khoảng thời gian bảy ngày ngay sau khi dữ liệu giao dịch hai năm được cung cấp.
- Mục đích của dự án xây dựng **1 two-system** để giới thiệu các mặt hàng của H&M cho cả khách hàng hiện tại và khách hàng mới đang tìm kiếm các mặt hàng tương tự.
- Đối với khách hàng hiện tại, **cách tiếp cận collaborative** được thực hiện bằng cách so sánh các hồ sơ khách hàng tương tự dựa trên các articles mà họ đã mua trong quá khứ.
- **Cách tiếp cận content-based**: đề xuất các bài viết dựa trên dữ liệu mô tả bài viết tương tự.



The screenshot shows the Kaggle competition page for "H&M Personalized Fashion Recommendations". The header features a banner with a row of wooden hangers and the text "Featured Prediction Competition", "H&M Personalized Fashion Recommendations", "Provide product recommendations based on previous purchases", and "\$50,000 Prize Money". Below the banner, it says "H&M Group · 2,952 teams · 9 months ago". The navigation bar includes "Overview", "Data", "Code", "Discussion", "Leaderboard", "Rules", "Team", "Submissions", and a "Late Submission" button. The "Overview" section is expanded, showing a table with the following content:

Overview	
Description	H&M Group is a family of brands and businesses with 53 online markets and approximately 4,850 stores. Our online store offers shoppers an extensive selection of products to browse through. But with too many choices, customers might not quickly find what interests them or what they are looking for, and ultimately, they might not make a purchase. To enhance the shopping experience, product recommendations are key. More importantly, helping customers make the right choices also has a positive implications for sustainability, as it reduces returns, and thereby minimizes emissions from transportation.
Evaluation	
Timeline	
Prizes	

# Bộ dữ liệu

- **Bộ dữ liệu gồm 4 file, trong đó:**
  - **images:** thư mục hình ảnh tương ứng với mỗi *article\_id*, hình ảnh trong các thư mục con bắt đầu bằng ba chữ số đầu tiên của *article\_id*. Lưu ý, không phải tất cả các *article\_id* đều có hình ảnh tương ứng.
  - **articles.csv:** metadata cho từng *article\_id* có sẵn để mua
  - **customers.csv:** metadata cho từng customer\_id trong dataset
  - **sample\_submission.csv:** file submission theo đúng định dạng nộp
  - **transactions\_train.csv:** dữ liệu train, bao gồm các giao dịch mua của từng khách hàng cho mỗi ngày, cũng như thông tin bổ sung. Các hàng trùng lặp tương ứng với nhiều lần mua cùng một mặt hàng.

	Total Rows	Total Columns
articles	105542	25
customers	1371980	7
transactions	31788324	5

# H & M Fashion (EDA)

prod_name	count
Timeless Midrise Brief	74662
Jade HW Skinny Denim TRS	68328
Tilly (1)	56688
Pluto RW slacks (1)	45625
Luna skinny RW	39466
Shake it in Balconette	39460
Primo slacks	39200
Tilda tank	39194
Cat Tee.	39077
Simple as That Triangle Top	35993

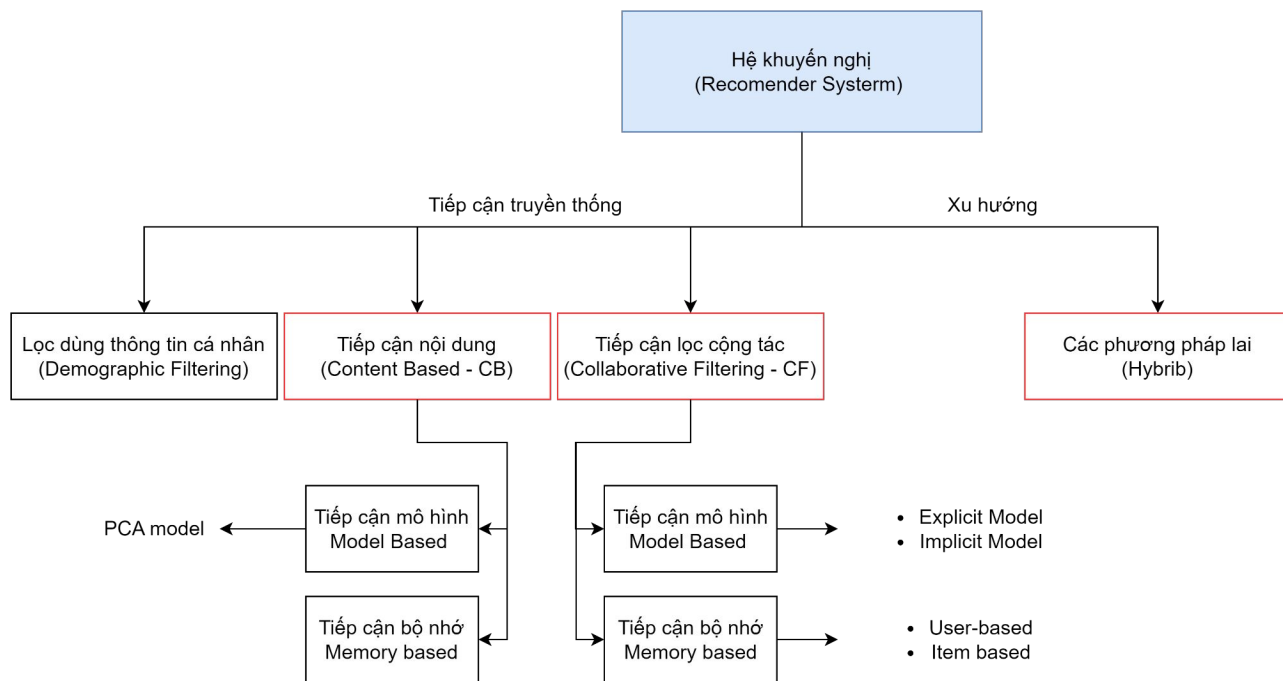
**Tóm tắt thống kê tên 1 số sản phẩm**

summary	price	summary	age	product_group_name	count
count	10980132	count	10940434	Garment Upper body	3942299
mean	0.027155150482678525	mean	35.21693289315579	Garment Lower body	2446417
stddev	0.017326580752706763	stddev	13.014627808510072	Garment Full body	1384796
min	3.389830508474576E-5	min	16	Swimwear	1094544
max	0.5067796610169492	max	99	Underwear	973634
colour_group_name	count	fashion_news_frequency	count	Accessories	507147
Black	3628137	NONE	6202387	Shoes	241317
White	1292709	Regularly	4752025	Socks & Tights	204275
Dark Blue	529338	null	24043	Nightwear	106430
Light Beige	511924	Monthly	1677	Unknown	71023
Blue	417358				
Beige	399221				
Light Pink	353195				
Light Blue	347879				
Off White	333948				
Greenish Khaki	314209				

only showing top 10 rows

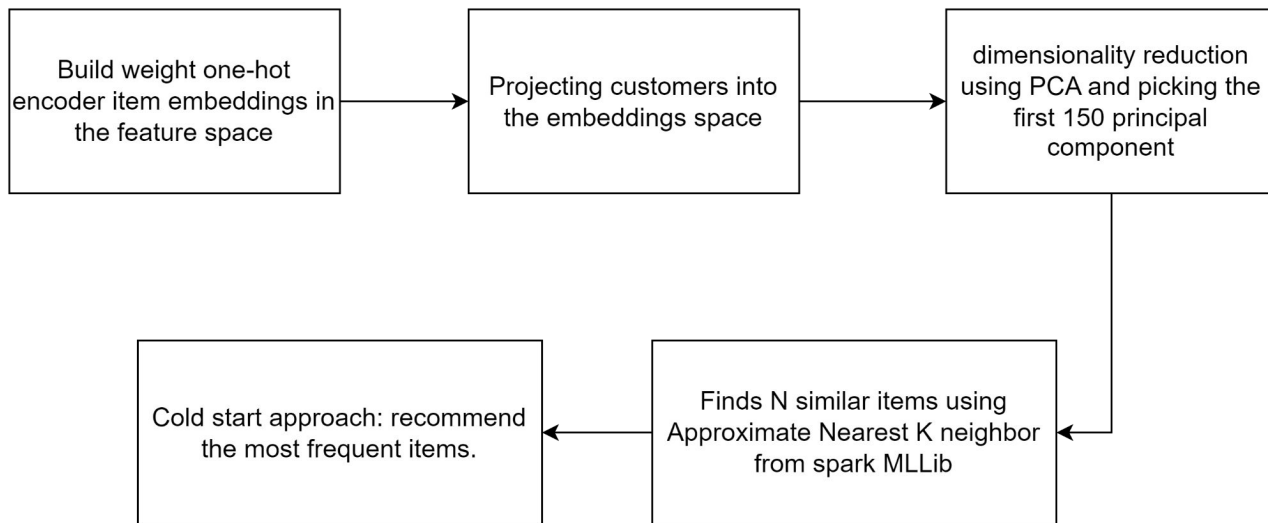
**Tóm tắt thống kê tên 1 số biến predict**

# Hướng tiếp cận



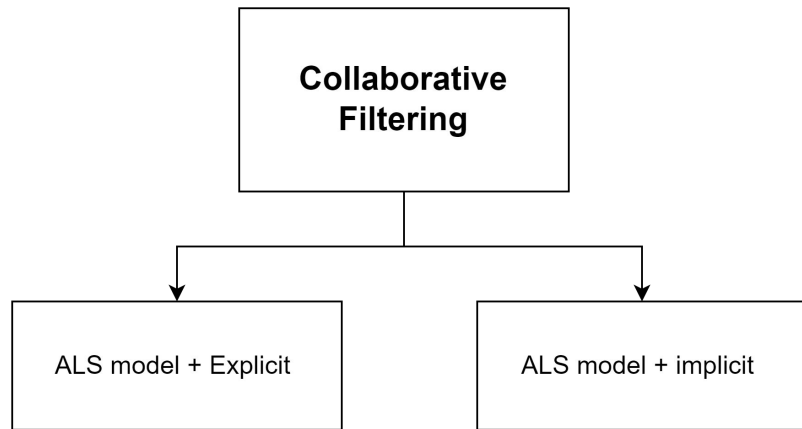
***Tổng quan phương pháp cho hệ khuyến nghị***

# Content-based Filtering

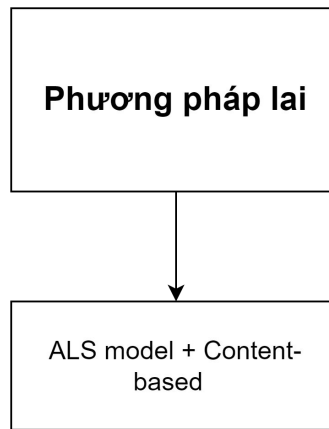


***Tổng quan phương pháp sử dụng PCA (content based filtering)***

# Collaborative Filtering + Phương pháp lai (Hybrid)



***Các phương pháp Collaborative Filtering***



***Phương pháp lai***

- **Explicit:** dựa trên hệ số ma trận coi các mục trong ma trận mục người dùng là tùy chọn rõ ràng do người dùng đưa ra đối với mục, ví dụ: người dùng đánh giá món hàng.
- **Implicit:** coi dữ liệu là những con số thể hiện cường độ quan sát hành động của người dùng (chẳng hạn như số lần nhấp vào sản phẩm). Những con số này sau đó liên quan đến mức độ tin cậy trong sở thích của người dùng, thay vì xếp hạng rõ ràng cho các mặt hàng. Sau đó, mô hình sẽ cố gắng tìm các yếu tố tiềm ẩn có thể được sử dụng để dự đoán sở thích dự kiến của người dùng đối với một mặt hàng.

# Kết quả

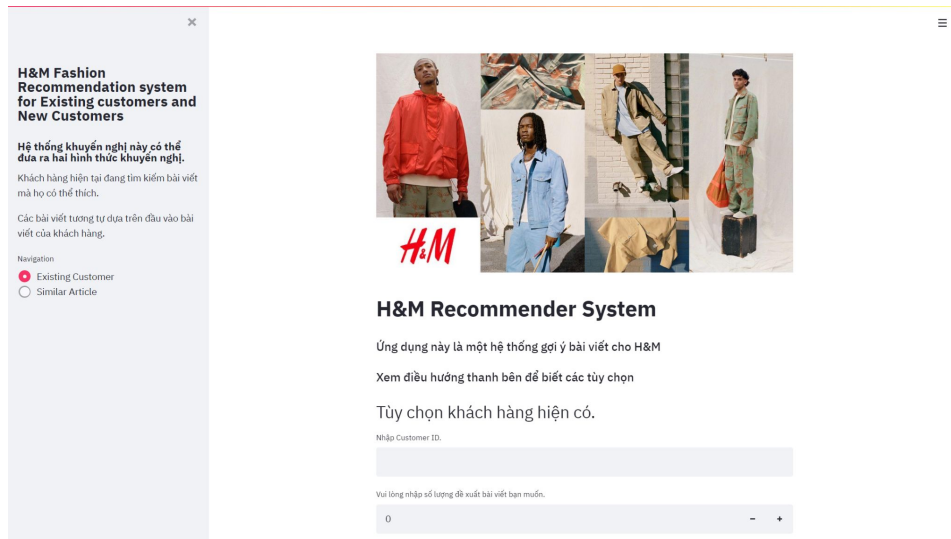
STT	Phương pháp	RMSE
1	Content-based using PCA model	0.82173
2	ALS model	0.51436
3	ALS model + Explicit	0.48625
4	ALS model + Implicit	1.24537
<b>5</b>	<b>ALS model + content-based</b>	<b>0.47076</b>

*Kết quả thực nghiệm*



# Kết luận

- Phương pháp cho kết quả tốt nhất là phương pháp lai (hybrid) kết hợp giữa lọc cộng tác (ALS model) và lọc dựa trên nội dung (Content-filtering) đạt RMSE = 0.47076.
- Hệ thống của chúng tôi có thể gợi ý sản phẩm cho khách hàng mới hoặc gợi ý sản phẩm cho khách hàng cũ.
- Để cải thiện hệ thống của chúng tôi, chúng tôi sẽ tiếp cận theo phương pháp Retrieval-Ranking.



***Hệ thống khuyến nghị thời trang H&M cho khách hàng hiện tại và khách hàng mới***