

Xây dựng hệ thống gợi ý sản phẩm cá nhân

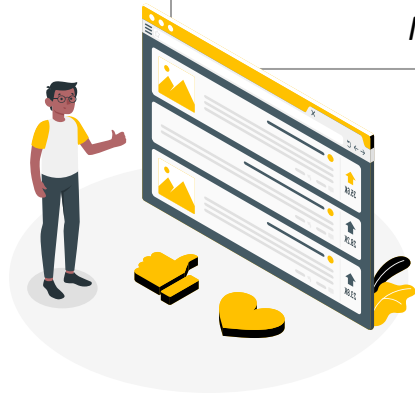
Nhóm 2

GVHD: TS. Nguyễn Tuấn Long



Thành viên

<i>Nguyễn Trần Tân An</i>	<i>12240001</i>
<i>Phạm Hồng Anh</i>	<i>11220534</i>
<i>Nguyễn Hải Đăng</i>	<i>12230014</i>
<i>Đỗ Ngọc Mai</i>	<i>11224014</i>
<i>Nguyễn Ngọc Tùng</i>	<i>12230075</i>



Nội dung

1. Giới thiệu đề tài
2. Tiền xử lý và khám phá dữ liệu
3. Xây dựng hệ thống gợi ý
4. Kết quả và đánh giá



Giới thiệu đề tài

01



1. Giới thiệu đề tài

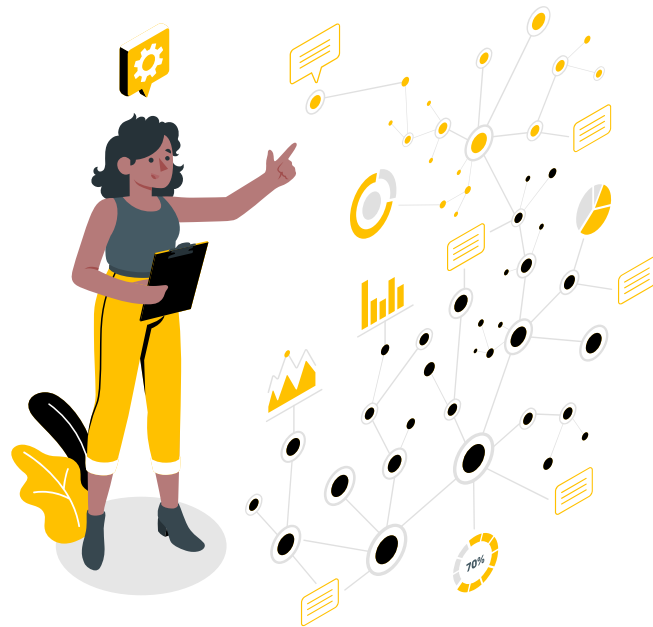
Trong quá trình xây dựng hệ thống gợi ý, nhóm đã cân nhắc hai hướng tiếp cận chính:

- User-based filtering – Dựa trên sự tương đồng giữa các khách hàng
- Item-based filtering – Dựa trên sự tương đồng giữa các sản phẩm

Mục tiêu của đề tài không chỉ dừng lại ở việc gợi ý sản phẩm, mà còn nhằm hiểu rõ hành vi mua hàng, khám phá các xu hướng nổi bật trong dữ liệu, từ đó đề xuất chiến lược tiếp thị phù hợp.

Tiền xử lý & khám phá dữ liệu

02



2. Tiền xử lý và khám phá dữ liệu

- Dữ liệu đầu vào gồm 203,422 dòng dữ liệu, mô tả chi tiết hóa đơn sản phẩm của một hệ thống thương mại điện tử
- Các cột dữ liệu:

Tên cột	Mô tả
InvoiceNo	Mã số hóa đơn
StockCode	Mã sản phẩm duy nhất đại diện cho từng mặt hàng
Description	Tên mô tả chi tiết sản phẩm
Quantity	Số lượng sản phẩm được mua trong hóa đơn
InvoiceDate	Ngày và giờ giao dịch được thực hiện
UnitPrice	Đơn giá của mỗi mặt hàng
CustomerID	Mã số khách hàng
Country	Quốc gia của khách hàng

2. Tiền xử lý và khám phá dữ liệu

Bước 1: Đọc dữ liệu và xử lý chuỗi thời gian:

- Tập trung phân tích giai đoạn từ tháng 2 đến tháng 6, loại bỏ các đơn hàng có số lượng âm hoặc 0, tạo biến Order Value để đo tổng chi tiêu mỗi đơn hàng.

Bước 2: Tóm tắt hóa đơn theo đơn vị giao dịch

- Gộp danh sách “StockCode” và “Description” theo “InvoiceNo”.
- Tính tổng giá trị hoá đơn “TotalOrderValue” và số mặt hàng “ItemCount”.
- Gộp tất cả thông tin thành “invoice_summary” làm đầu vào phân tích.

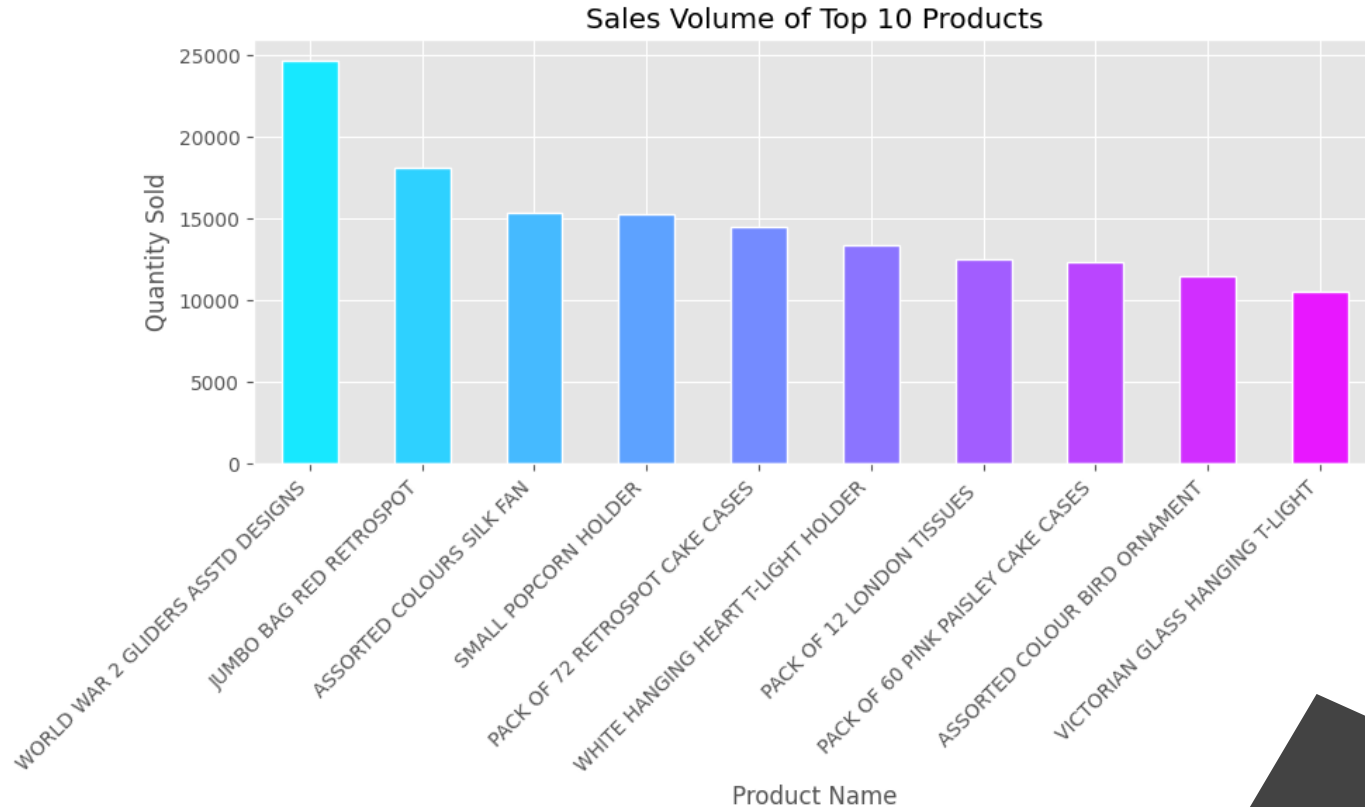
Bước 3: Phân tích số lượng bán theo sản phẩm

- Xác định các sản phẩm bán chạy nhất

Bước 4: Tính tần suất và giá trị mua hàng trung bình

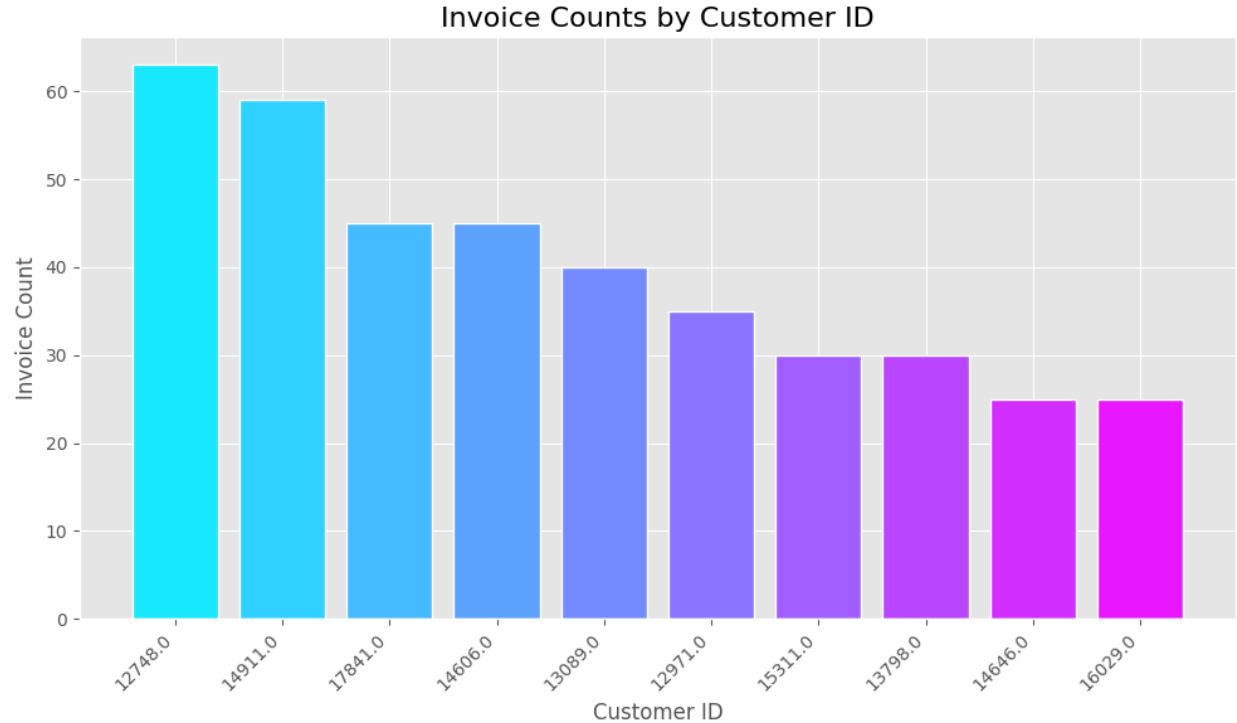
- Đếm số lần mỗi CustomerID xuất hiện
- Tính giá trị mua hàng trung bình

2. Tiền xử lý và khám phá dữ liệu



2. Tiền xử lý và khám phá dữ liệu

Top 10 CustomerID có tần suất mua nhiều nhất



Xây dựng Hệ thống gợi ý

03



3. Xây dựng hệ thống gợi ý

Tạo ma trận Customer-Item matrix

```
customer_item_matrix = df.pivot_table(  
    index='CustomerID',  
    columns='StockCode',  
    values='Quantity',  
    aggfunc='sum'  
)
```



3. Xây dựng hệ thống gợi ý

Phần 1: User-based filtering

Bước 1: Tính toán độ tương đồng giữa người dùng (sử dụng hàm `cosine_similarity`)

```
user_user_sim_matrix = pd.DataFrame(  
    cosine_similarity(customer_item_matrix))
```

Kết quả là một ma trận vuông, trong đó:

- Hàng và cột là CustomerID
- Mỗi ô là chỉ số tương đồng từ 0 → 1
- Càng gần 1 thì càng giống nhau về hành vi mua hàng

3. Xây dựng hệ thống gợi ý

Phần 1: User-based filtering

Bước 2: Xây dựng hệ thống gợi ý cho user-based

Chọn khách hàng mục tiêu A, ví dụ CustomerID=15240

Tìm các khách hàng B có độ tương đồng cao nhất với người đó

```
user_user_sim_matrix.loc[15240.0].sort_values(ascending=False)
```

Thu được kết quả, CustomerID=16407 có tương quan lớn nhất tới 15240

Tìm các sản phẩm khách hàng B đã mua nhưng khách hàng A chưa mua để đưa ra gợi ý

3. Xây dựng hệ thống gợi ý

Phần 1: User-based filtering

Bước 2: Xây dựng hệ thống gợi ý cho user-based

Tạo danh sách sản phẩm đề xuất cho khách hàng A

```
df.loc[
    df['StockCode'].isin(items_to_recommend_to_A),
    # lọc ra StockCode
    ['StockCode', 'Description']
    # Lấy 2 cột StockCode và Description
].drop_duplicates().set_index('StockCode')
```

Ta thu được 1 danh sách gồm 27 sản phẩm được gợi ý.

3. Xây dựng hệ thống gợi ý

Phần 2: Item-based filtering

Bước 1: Tính toán độ tương đồng giữa sản phẩm - sản phẩm

```
item_item_sim_matrix ==  
pd.DataFrame(cosine_similarity(customer_item_matrix.T),  
index=customer_item_matrix.columns,  
columns=customer_item_matrix.columns)
```

Kết quả là một ma trận vuông, trong đó:

- Hàng và cột là StockCode
- Mỗi ô là chỉ số tương đồng từ 0 \rightarrow 1
- Càng gần 1 thì các mặt hàng càng khả năng tiêu thụ giống nhau

3. Xây dựng hệ thống gợi ý

Phần 2: Item-based filtering

Bước 2: Xây dựng hệ thống gợi ý cho item-based

Chọn sản phẩm mục tiêu A, ví dụ StockCode 22708

Tìm 10 sản phẩm giống sản phẩm StockCode 22708 nhất

```
top_10_similar_items = list(  
    item_item_sim_matrix.loc["22708"]  
    .sort_values(ascending=False)  
    .iloc[:11]  
    .index  
)
```

Ta thu được danh sách 10 sản phẩm tương đồng với A.

3. Xây dựng hệ thống gợi ý

Top 10 sản phẩm tương đồng
với StockCode 22708

StockCode	Description
22708	WRAP DOLLY GIRL
22308	TEA COSY BLUE STRIPE
85071C	CHARLIE+LOLA"EXTREMELY BUSY" SIGN
21902	KEY FOB , FRONT DOOR
23176	ABC TREASURE BOOK BOX
22981	PANTRY APPLE CORER
21899	KEY FOB , GARAGE DESIGN
21900	KEY FOB , SHED
22615	PACK OF 12 CIRCUS PARADE TISSUES
23242	TREASURE TIN BUFFALO BILL
85071A	BLUE CHARLIE+LOLA PERSONAL DOORSIGN

Kết quả & đánh giá

04



4. Kết quả và đánh giá

Kết quả:

- Xây dựng thành công hệ thống gợi ý theo item-based và user-based
- Hiểu được cách một số hệ thống gợi ý phân tích và xử lý dữ liệu



4. Kết quả và đánh giá

Đề xuất phát triển:

- Sử dụng hybrid recommender (content + collaborative)
- Tích hợp vào hệ thống engine gợi ý của các nền tảng thương mại điện tử



Cảm ơn thầy và
các bạn đã lắng
nghe

