

SAMSUNG

I-Powered Product Recommendation System for E-commerce

OU _ BD2 _ GROUP 3

| Big Data Course

I-Powered Product Recommendation System for E-commerce

OU _ BD2 _ GROUP 3

Big Data Course

• THÀNH VIÊN NHÓM

- Hồ Ngọc Nhung
- Bùi Ngọc Phương Linh
- Phạm Thị Khánh Ly
- Bùi Dạ Lý
- Yoon Min
- Phan Lê Nguyên
- Trần Nhật Minh



NỘI DUNG

| 1. Giới thiệu dự án


| 2. Quy trình thực hiện dự án


| 3. Cải tiến trong tương lai

| 4. Kết quả



Dự án này xuất phát từ nhu cầu thực tế trong việc cải thiện trải nghiệm người dùng thông qua các gợi ý sản phẩm cá nhân hóa.





Size Chart

Product details

Fabric type 100% Leather

Care instructions Machine Wash

Origin Imported


Sole material Rubber

About this item


- Leather upper
- ABZORB midsole cushioning aids in absorbing forceful impacts
- Premium PU comfort insert

Products related to this item


Sponsored




Lugz Men's Clipper Sneaker, White, 10 D US
★★★★★ 32,822
\$34⁹⁵
prime




Mizuno Men's Wave Legacy Sneaker, White, 10 Wide
★★★★★ 148
\$102⁵³
prime




ECCO Men's Biom 2.2 Textile Cross Trainer, Acorn/Black, 10-10.5
★★★★★ 4
\$126²⁶
prime




ECCO Men's Biom 2.2 Sport Cross Trainer, Black/Limestone/Sand, 11-11.5
★★★★★ 8
\$133⁹⁶
prime



ECCO Men's Biom 2.2 Cross Trainer, Sand Nubuck, 11-11.5
★★★★★ 12
\$179⁰²
prime
Prime Try Before You Buy





Size Chart

Product details

Fabric type 100% Leather

Care instructions Machine Wash

Origin Imported

Sole material Rubber


About this item

- Leather upper
- ABZORB midsole cushioning aids in absorbing forceful impacts
- Premium PU comfort insert


Similar items

Sponsored

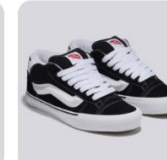
[Feedback on our suggestions](#) | [See all](#)




Vans Knu Skool Skate Shoes Sneakers VN0009QC6BT Black True White



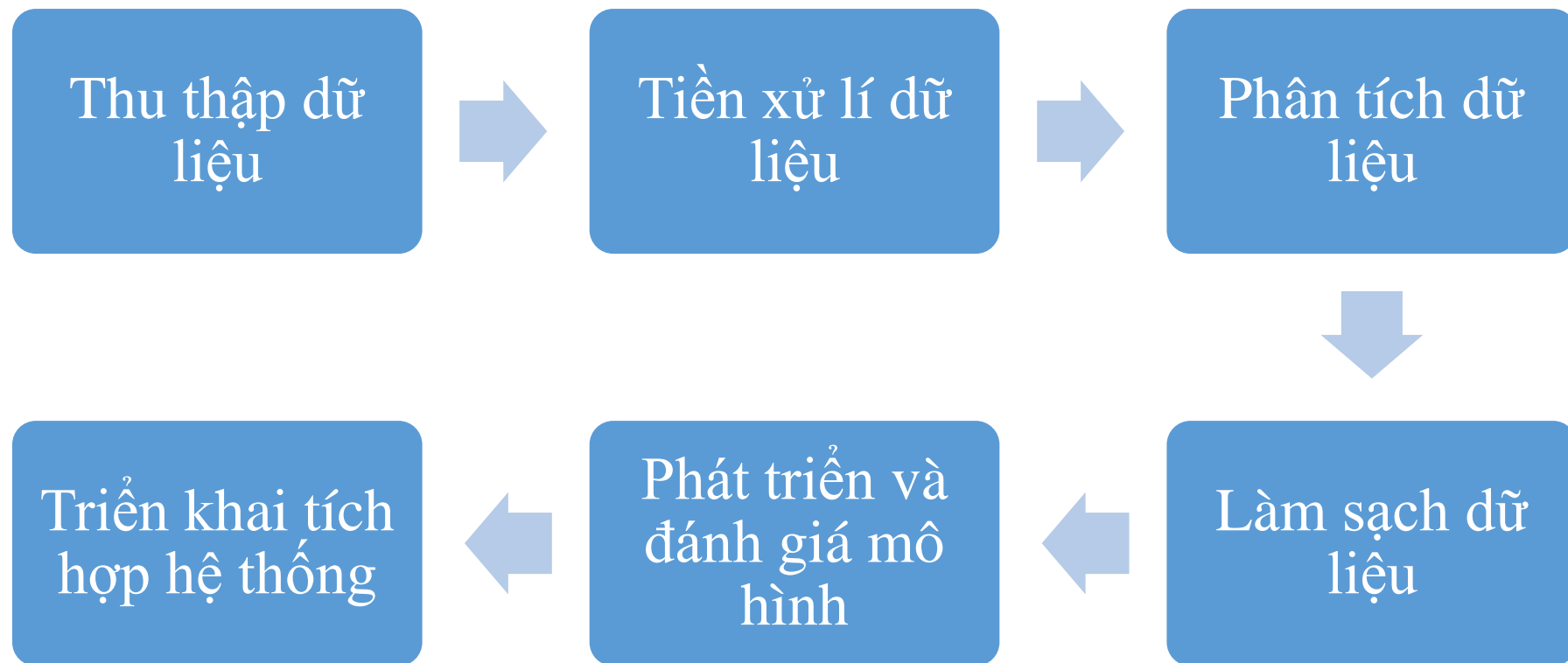
Vans Knu Skool Skate Shoes Sneakers Black/True White VN0009QC6BT US 4-13



Vans Knu Mid Skate Shoes Sneakers Fast Shipping VN000CQ96BT1



VANS Knu Skool OG 2023 MEN'S SNEAKERS VN0009QC85P BLACK...



Giai đoạn 1: Thu thập dữ liệu

- I Dữ liệu được lấy từ dữ liệu công khai trên nền tảng github
- I Dữ liệu ban đầu có:
 - ▶ 32 thuộc tính: Uniq Id, Crawl Timestamp, Dataset Origin, Product Id, Product Barcode, Product Company Type Source, Product Brand Source, Product Brand Normalised Source,...
 - ▶ 5000 dòng dữ liệu
 - ▶ Dữ liệu thô chưa qua xử lí
 - ▶ Tình trạng: Nhiều thuộc tính dư thừa, nhiều giá trị null, chưa được chuẩn hóa,...

Uniq Id ▾	Crawl Timestamp ▾	Dataset Origin ▾	Product Id ▾	Product Barcode ▾	Product Company Type Source ▾	Product Brand Source ▾
1705736792d82aa2f2d3caf1c07c53f4	2020-09-24 03:21:12 +0000	<null>	2e17bf4acecdece67fc00f07ad62c910	<null>	Competitor	<null>
95a9fe6f4810fcfc7ff244fd06784f11	2020-10-30 14:04:08 +0000	<null>	076e5854a62dd283c253d6bae415af1f	<null>	Competitor	<null>
8d4d0330178d3ed181b15a4102b287f2	2020-08-06 05:51:47 +0000	<null>	8a4fe5d9c7a6ed26cc44d785a454b124	<null>	Competitor	<null>
fdcdc4df45b35efd886794b261f730c51	2020-07-15 11:22:04 +0000	<null>	03b5fb878a33eadff8b033419eab9669	<null>	Competitor	
0990cf89a59ca6a0460349a3e4f51d42	2020-11-26T12:27:20+00:00	<null>	ce3d761e57d6ccad80619297b5b1bcbc	<null>	Competitor	
262bcd1c9a561f6c6ad12bfdddc0d99e	2020-09-29 21:42:41 +0000	<null>	7569141059e839b828a5bd3887243c1a	<null>	Competitor	
05224642381f73764e85a3382ee0ff61	2020-07-29 19:12:55 +0000	<null>	b371dbac1c2d9457f61742c3c39bbdf3	<null>	Competitor	<null>

Giai đoạn 2: Tiền xử lý dữ liệu

- 1. Lọc dữ liệu để chỉ giữ lại các cột cần thiết:

```
train_data = train_data[['Uniq Id', 'Product Id', 'Product Rating', 'Product Reviews Count',  
                          'Product Category', 'Product Brand', 'Product Name',  
                          'Product Image Url', 'Product Description', 'Product Tags']]  
train_data.head(7)
```

- 2. Xử lý giá trị thiếu: Thay thế hoặc loại bỏ các giá trị bị thiếu để đảm bảo tính đầy đủ của dữ liệu.

```
train_data.isnull().sum()  
  
train_data['Product Rating'].fillna(0, inplace=True)  
train_data['Product Reviews Count'].fillna(0, inplace=True)  
train_data['Product Category'].fillna('', inplace=True)  
train_data['Product Brand'].fillna('', inplace=True)  
train_data['Product Description'].fillna('', inplace=True)  
train_data.isnull().sum()
```

Giai đoạn 2: Tiền xử lý dữ liệu

- Xử lý giá trị thiếu:** Thay thế hoặc loại bỏ các giá trị bị thiếu để đảm bảo tính đầy đủ của dữ liệu.

```
train_data.isnull().sum()
```

	0
Uniq Id	0
Product Id	0
Product Rating	2806
Product Reviews Count	1654
Product Category	10
Product Brand	13
Product Name	0
Product Image Url	0
Product Description	1127
Product Tags	0



```
train_data.isnull().sum()
```

	0
Uniq Id	0
Product Id	0
Product Rating	0
Product Reviews Count	0
Product Category	0
Product Brand	0
Product Name	0
Product Image Url	0
Product Description	0
Product Tags	0

Giai đoạn 2: Tiền xử lý dữ liệu

- Loại bỏ bản ghi trùng lặp:** Xóa các bản ghi trùng lặp để tránh làm sai lệch kết quả phân tích.

```
train_data.duplicated().sum()
```

- Chuyển đổi kiểu dữ liệu:** Đảm bảo rằng các cột có kiểu dữ liệu chính xác, ví dụ: chuyển đổi chuỗi số thành số thực.

```
column_name_mapping = {
    'Uniq Id': 'ID',
    'Product Id': 'ProdID',
    'Product Rating': 'Rating',
    'Product Reviews Count': 'ReviewCount',
    'Product Category': 'Category',
    'Product Brand': 'Brand',
    'Product Name': 'Name',
    'Product Image Url': 'ImageURL',
    'Product Description': 'Description',
    'Product Tags': 'Tags',
    'Product Contents': 'Contents'
}

train_data.rename(columns=column_name_mapping, inplace=True)
train_data['ID'] = train_data['ID'].str.extract(r'(\d+)').astype(float)
train_data['ProdID'] = train_data['ProdID'].str.extract(r'(\d+)').astype(float)
train_data.head(3)
```

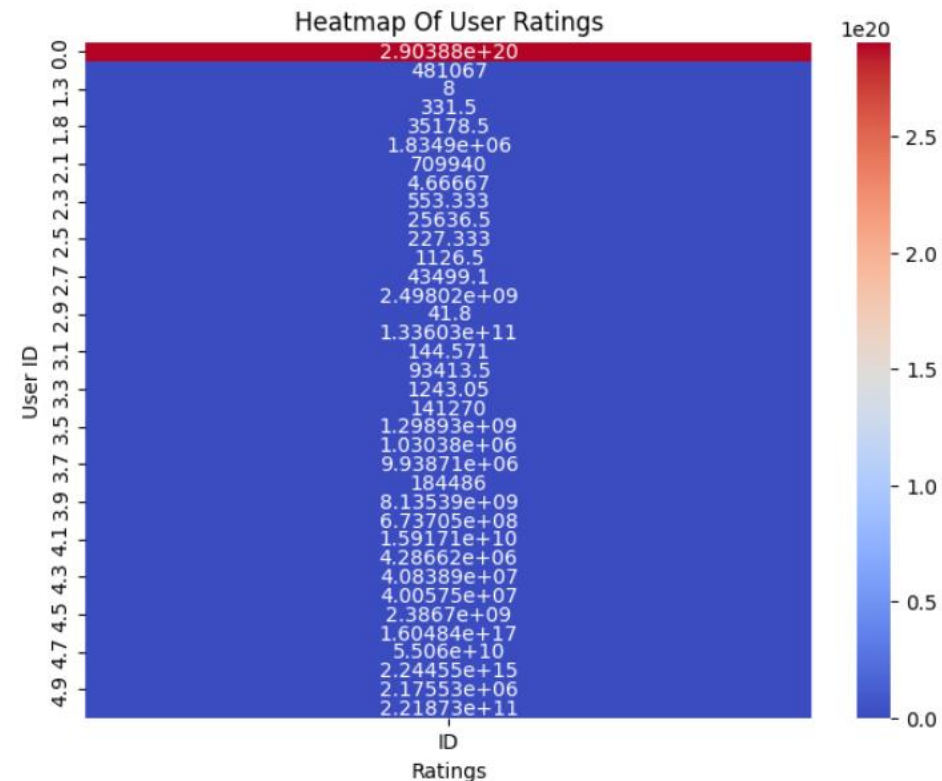
Giai đoạn 3: Phân tích dữ liệu

I Phân Tích Số Lượng Người Dùng, Sản Phẩm, và Đánh Giá:

► Mục đích: Đếm số lượng người dùng, sản phẩm và mức đánh giá duy nhất trong dữ liệu. Điều này cung cấp cái nhìn tổng quan về kích thước và độ phân bố của các yếu tố trong bộ dữ liệu..

```
num_users = train_data['ID'].nunique()
num_items = train_data['ProdID'].nunique()
num_ratings = train_data['Rating'].nunique()
print(f"Number of unique users: {num_users}")
print(f"Number of unique items: {num_items}")
print(f"Number of unique ratings: {num_ratings}")

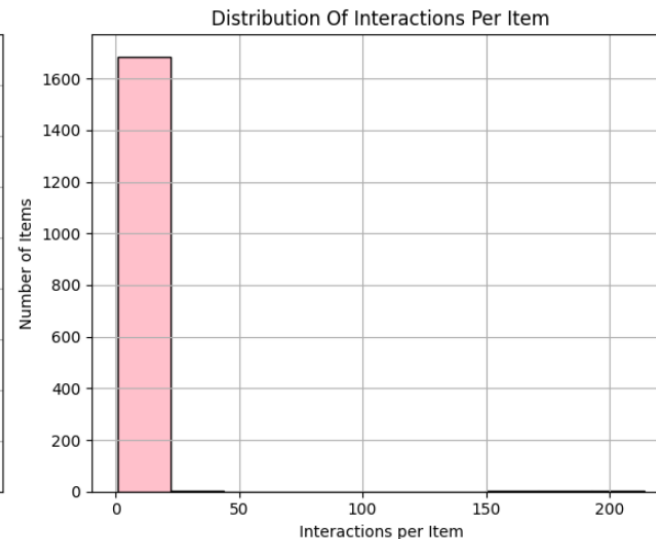
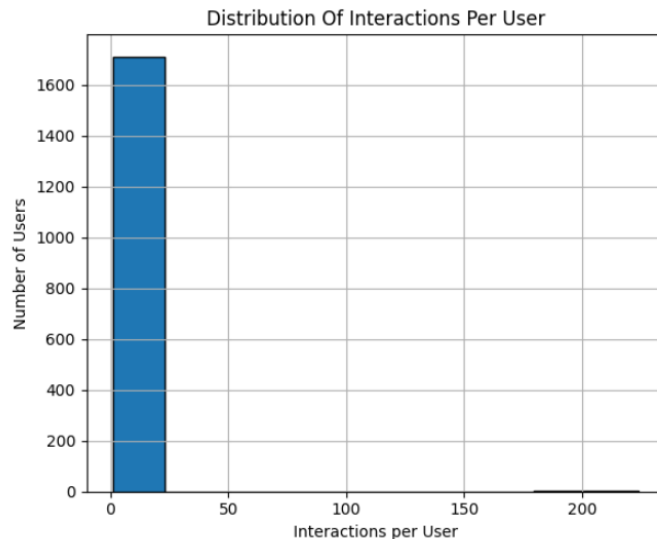
#heatmap
heatmap_data = train_data.pivot_table('ID', 'Rating')
```



Giai đoạn 3: Phân tích dữ liệu

I Phân Tích Phân Phối Tương Tác:

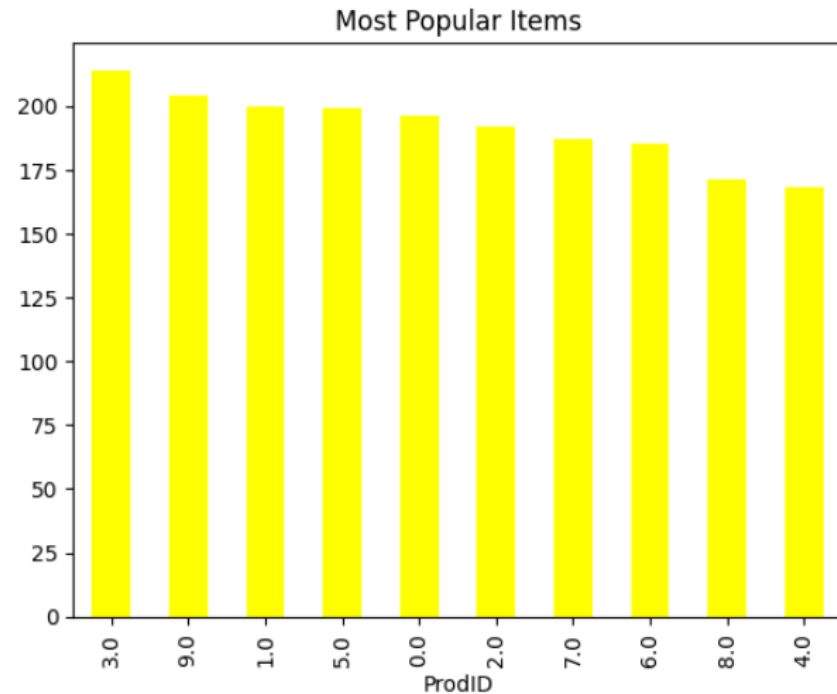
- ▶ Giúp nhận diện các mẫu trong mức độ hoạt động của người dùng và sản phẩm.



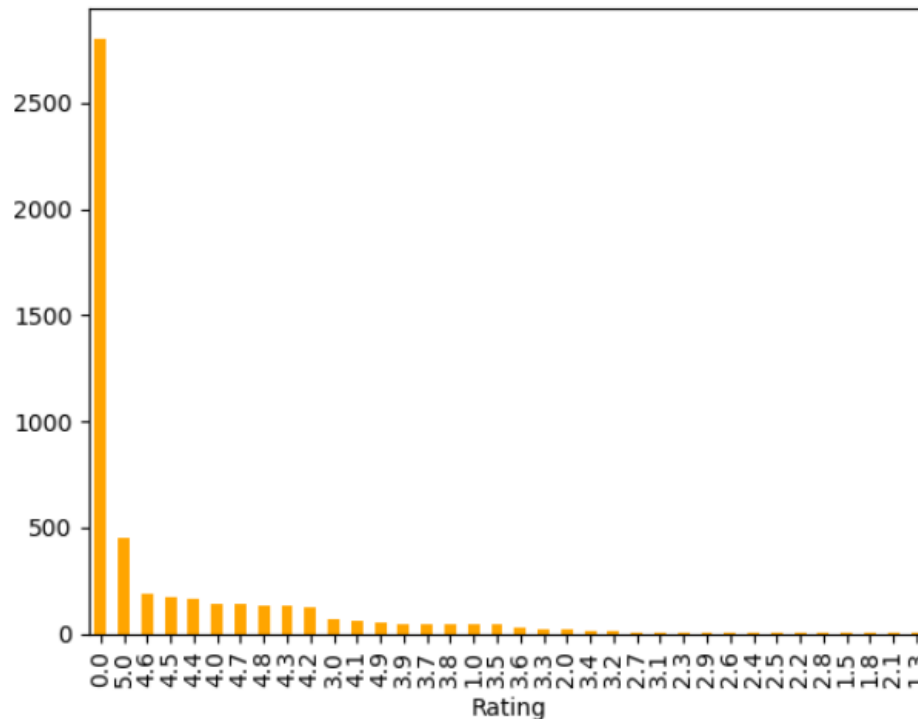
Giai đoạn 3: Phân tích dữ liệu

I Hiển Thị Các Sản Phẩm Phổ Biến:

- Mục đích: Hiển thị 10 sản phẩm phổ biến nhất dựa trên số lượng tương tác. Cung cấp cái nhìn về các sản phẩm được ưa chuộng trong dữ liệu.



Giai đoạn 3: Phân tích dữ liệu



I Hiện Thị Các Mức Đánh Giá Phổ Biến:

- ▶ Mục đích: Hiện thị phân phối các mức đánh giá trong dữ liệu. Cung cấp cái nhìn về cách đánh giá phân bố giữa các mức độ khác nhau.

Giai đoạn 4: Làm sạch dữ liệu

I Xử lý và Tinh chỉnh Tags:

- ▶ Chuẩn hóa các dữ liệu văn bản
- ▶ Chuyển đổi văn bản thành chữ thường, loại bỏ các từ dừng (stop words) và chỉ giữ lại các từ có nghĩa (chứa ký tự chữ và số).
- ▶ Loại bỏ từ không cần thiết và chỉ giữ lại các từ quan trọng.
- ▶ Tạo cột Tags mới

Tags	
0	OPI Infinite Shine, Nail Lacquer Nail Polish, ...
1	Nice 'n Easy Permanent Color, 111 Natural Medi...
2	Clairol Nice 'N Easy Permanent Color 7/106A Na...



Tags	
0	premium, beauty, premium, makeup, premium, nai...
1	beauty, hair, care, hair, color, auburn, hair,...
2	beauty, hair, care, hair, color, permanent, ha...

Giai đoạn 5: Phát triển và đánh giá mô hình

| Các mô hình thử nghiệm:

- ▶ K-Means Clustering
- ▶ K-Nearest Neighbors (KNN)
- ▶ Random Forest
- ▶ Support Vector Machines (SVM)
- ▶ Logistic Regression
- ▶ Decision Tree
- ▶ Neural Network

Giai đoạn 5: Phát triển và đánh giá mô hình

I K-Nearest Neighbors (KNN)

- ▶ Mục tiêu: Phân loại sản phẩm vào các nhóm đã phân cụm bằng K-Means.
- ▶ Quy trình:
 - Chọn các đặc trưng đã mã hóa để huấn luyện mô hình.
 - Áp dụng StandardScaler để chuẩn hóa dữ liệu.
 - Sử dụng KNN với $k=4$ để dự đoán nhóm sản phẩm.
 - Đánh giá mô hình bằng độ chính xác (accuracy).
 - Với $k = 6$ ta có độ chính xác là 55.20%

Accuracy: 55.20%

Giai đoạn 5: Phát triển và đánh giá mô hình

I Random Forest Classifier

- ▶ Mục tiêu: Phân loại sản phẩm dựa trên các đặc trưng mã hóa.
- ▶ Quy trình:
 - Chọn các đặc trưng tương tự như trong KNN.
 - Áp dụng Random Forest với 1000 cây quyết định để dự đoán nhóm sản phẩm.
 - Đánh giá mô hình bằng độ chính xác (accuracy).

```
# Chuẩn hóa dữ liệu
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

model = RandomForestClassifier(n_estimators=1000, random_state=42)

# Huấn luyện mô hình
model.fit(X_train, y_train)

# Dự đoán
y_pred = model.predict(X_test)

# Đánh giá mô hình
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy* 100:.2f}%')
```

Accuracy: 79.50%

Giai đoạn 5: Phát triển và đánh giá mô hình

| Logistic Regression

- ▶ Mục tiêu: Phân loại sản phẩm và so sánh với các mô hình khác.
- ▶ Quy trình:
 - Áp dụng Logistic Regression với 10000 lần lặp.
 - Đánh giá mô hình bằng độ chính xác

Accuracy: 48.90%

```
# Chuẩn hóa dữ liệu
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Khởi tạo mô hình Logistic Regression
model = LogisticRegression(max_iter=10000, random_state=42)

# Huấn luyện mô hình
model.fit(X_train, y_train)

# Dự đoán
y_pred = model.predict(X_test)

# Đánh giá mô hình
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy* 100:.2f}%')
```

Giai đoạn 6: Triển khai tích hợp hệ thống

I Ứng dụng Flask Web

- ▶ Hệ thống được triển khai dưới dạng ứng dụng web sử dụng Flask.
- ▶ Giao diện web đơn giản cho phép người dùng nhập tên sản phẩm và nhận gợi ý sản phẩm tương tự.

Product Suggestion System

Enter the name of a product to get similar product suggestions.

Giai đoạn 6: Triển khai tích hợp hệ thống

I Tích hợp và Kết nối











- ▶ Mô hình và dữ liệu được kết nối với ứng dụng web, tạo ra quy trình dự đoán liên tục.
- ▶ Sử dụng AJAX trong giao diện web để gửi yêu cầu dự đoán và cập nhật kết quả mà không cần tải lại trang.

I Gợi ý sản phẩm

- ▶ Hệ thống hiển thị tối đa 10 sản phẩm tương tự dựa trên dự đoán của mô hình.
- ▶ Kết quả bao gồm thông tin chi tiết về sản phẩm như tên, danh mục, thương hiệu và hình ảnh.


Giai đoạn 6: Triển khai tích hợp hệ thống

Get Suggestions

 <p>Sorry, photo not available</p> <p>BMC Bright and Loud Cream Gel Lacquer Polish Set - Neon Wasteland Collection</p> <p>Category: beauty, makeup, nails</p> <p>Brand: bmc</p>	 <p>(4 Pack) Pure Ice Base Coat Nail Polish, Bottoms Up, 0.5 fl oz</p> <p>Category: beauty, makeup, nails</p> <p>Brand: pure, ice</p>	 <p>Sorry, photo not available</p> <p>Sinful Colors Professional Nail Polish, Show & Teal, 0.5 fl oz</p> <p>Category: beauty, makeup, nails</p> <p>Brand: sinfulcolors</p>	 <p>Sorry, photo not available</p> <p>Sally Hansen Removers, Extra Strength Remover 8 Oz</p> <p>Category: beauty, makeup, nails</p> <p>Brand: sally, hansen</p>	 <p>Unique Bargains Nail File Buffer 10 Pcs 4 Way Shine Edge Removal Polish Manicure Tool Black Red</p> <p>Category: beauty, makeup, nails</p> <p>Brand: dl</p>	 <p>Sorry, photo not available</p> <p>essie Gel Couture 2-Step Longwear Nail Polish, Lace Me Up, Pink Nude Nail Polish, 0.46 fl. oz.</p> <p>Category: beauty, makeup, nails</p> <p>Brand: essie</p>
 <p>Sorry, photo not available</p> <p>NYC New York Color In a New York Color Minute Quick Dry Nail Polish, 224 Times Square, 0.33 fl oz</p>	 <p>Sorry, photo not available</p> <p>L'Oreal Paris Colour Riche Nail Color, 600 Douzens Nude, 0.39 fl oz</p>	 <p>OPI Nail Lacquer Polish .5oz/15mL - Sweet Heart S96</p> <p>Category: beauty, makeup, nails</p>	 <p>Kokie Professional Nail Polish, Girls Night, 0.54 fl oz</p> <p>Category: beauty, makeup, nails</p>		

Giai đoạn 6: Triển khai tích hợp hệ thống


Get Suggestions



Crest Complete Whitening + Scope Toothpaste, 5 Ct, 5.8 Oz

Category: personal, care, personal, care, brand, crest


Brand: crest



Crest 4.8 Oz 3D White Toothpaste

Category: personal, care, personal, care, brand, crest


Brand: crest



Crest Gum and Breath Purify Toothpaste, Deep Clean, 4.1 oz

Category: personal, care, personal, care, brand, crest

Brand: crest




Sorry, photo not available

(4 pack) Crest Pro-Health Sensitive & Enamel Shield Toothpaste, 4.6 oz

Category: personal, care, personal, care, brand, crest

Brand: crest




Sorry, photo not available

Crest 3D White, Whitening Toothpaste, Radiant Mint, 4.8 Ounce, Pack of 3 NEW

Category: personal, care, personal, care, brand, crest

Brand: crest




Sorry, photo not available

Crest 3D White Brilliance Mouthwash, Alcohol Free, Mint, 33.8 fl oz

Category: personal, care, personal, care, brand, crest


Brand: crest



Sorry, photo not available

Crest Pro-Health Sensitive & Enamel Shield Toothpaste, 4.6 oz (Pack of 2)


Category: personal, care, personal, care, brand, crest



Crest Kids Cavity Protection Toothpaste

Category: personal, care, personal, care, brand, crest


Brand: crest



6 Pack - 3D White Brilliance Toothpaste, Vibrant Peppermint 4.1 oz

Category: personal, care, personal, care, brand, crest

Brand: crest



Sorry, photo not available

Crest + Scope Complete Whitening Toothpaste, Minty Fresh, .85 oz

Category: personal, care, personal, care, brand, crest

Brand: crest

| Mở rộng dữ liệu

- ▶ Thêm dữ liệu sản phẩm mới để cải thiện độ chính xác của mô hình và mở rộng khả năng gợi ý.

| Tối ưu hóa mô hình

- ▶ Cải thiện và tinh chỉnh mô hình để tăng hiệu suất dự đoán và thời gian phản hồi.

| Xây dựng một hệ thống hoàn chỉnh

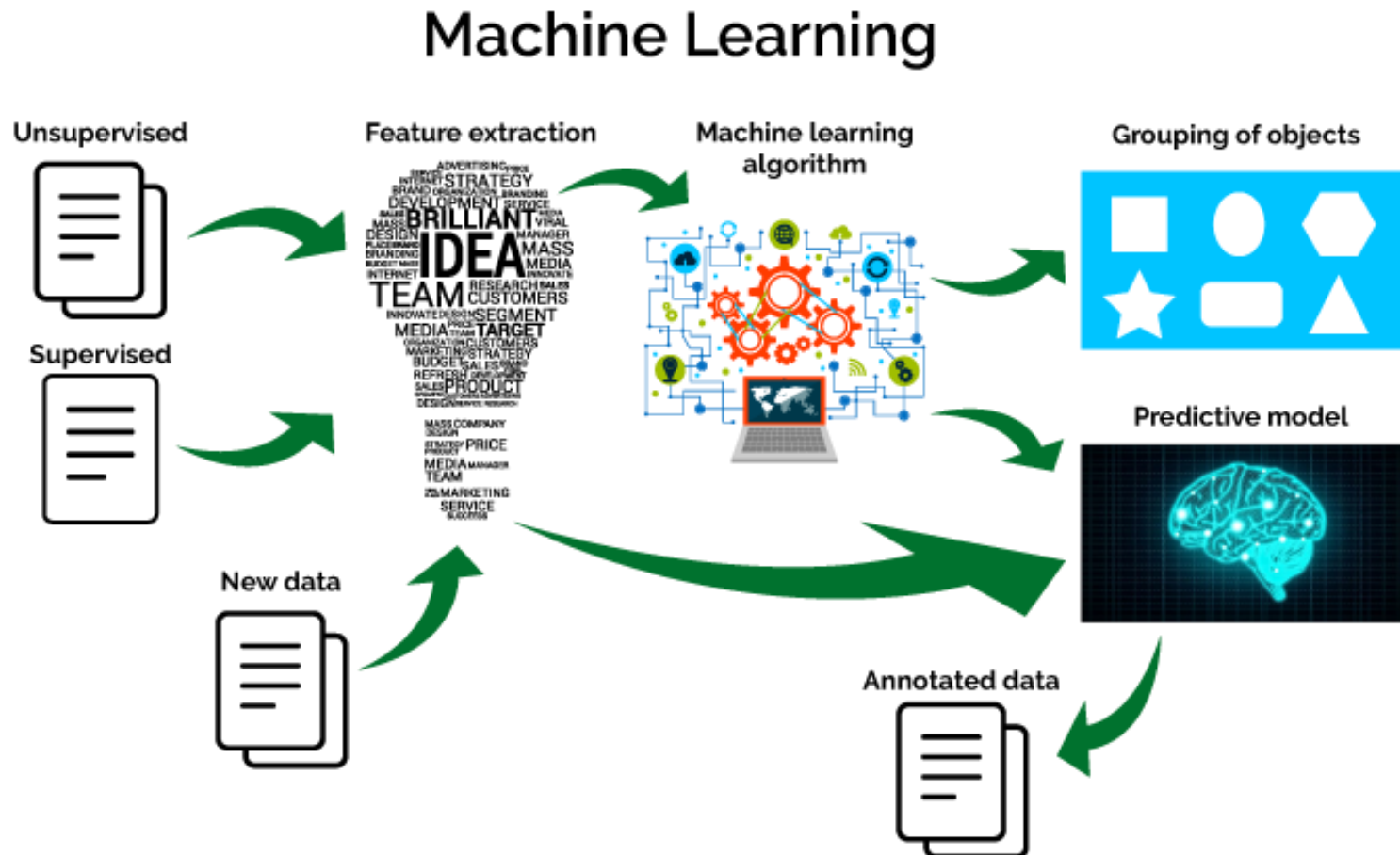
- ▶ Xây dựng website bán hàng trực tuyến tích hợp hệ thống gợi ý sản phẩm.

1. Kinh nghiệm làm việc với dữ liệu



2. Kỹ năng lập trình và tích hợp

3. Hiểu biết về mô hình máy học



4. Quản lý dự án và làm việc nhóm

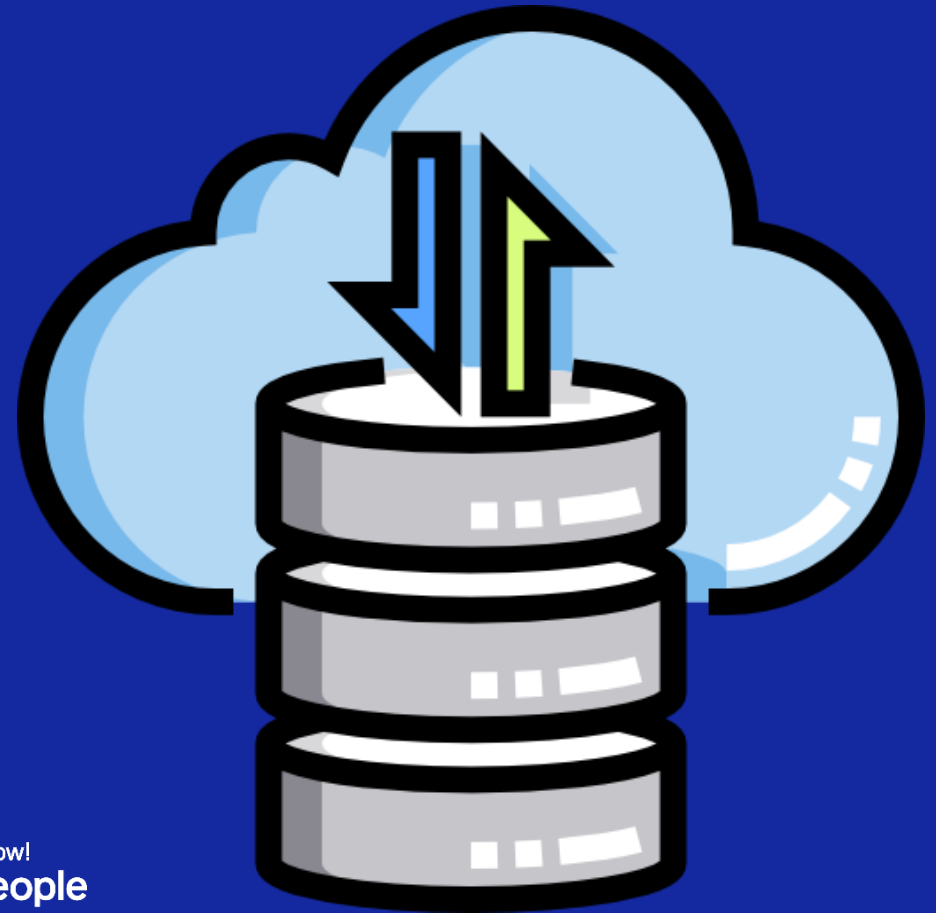


SAMSUNG

OU _ BD2 _ GROUP 3

I-Powered Product Recommendation System for E-commerce

**Cảm ơn thầy
và các bạn đã
lắng nghe!**





SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations

©2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung Innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.