

# eCommerce 구매전환율 제고를 위한 상품 추천 시스템 구현

15기 남연우

# 목차

01. 프로젝트 개요

02. 프로젝트 수행 절차 및 방법

03. 프로젝트 수행 결과

04. 자체 평가 의견

## 01. 프로젝트 개요

---

[1]마케팅 솔루션 기업 빅인의 '2021년 국내 이커머스 트렌드 리포트'에 따르면,  
온라인 쇼핑몰의 평균 구매 전환율은 **2.06%** 대에 이른다고 합니다.  
그에 비해 우리 온라인 쇼핑몰의 구매 전환율은 **1.57%**에 그치고 있습니다.

Objective : 전환율과 매출을 올리기 위해서 고객에게 적절한 상품을 제안하는  
**추천 시스템**을 구현하고자 합니다.

※ 본 프로젝트는 중동지역의 대규모 다중 카테고리 온라인 상점의 고객 행동 데이터를 분석한 보고서입니다.

---

[1]출처 : E-commerce Trend Report Vol.04 - 이커머스 전환의 모든 것, 빅인, 2021년 9월 작성,

[https://biginsight.io/trend-report-2021-conversion?utm\\_source=Di&utm\\_medium=btn&utm\\_campaign=trendreport\\_v4&utm\\_content=trendreport\\_v4&utm\\_term=Di\\_subscriber](https://biginsight.io/trend-report-2021-conversion?utm_source=Di&utm_medium=btn&utm_campaign=trendreport_v4&utm_content=trendreport_v4&utm_term=Di_subscriber)

# 01. 프로젝트 개요

- ▶ 데이터는 2019년 11월 한달 간의 이커머스 고객 행동에 대한 로그 데이터이며 제품 확인, 장바구니, 구매, 그리고 제품에 대한 정보로 약 6,700만개의 행이 구성되어 있습니다.

	event_time	event_type	product_id	category_code	sub_category	product	brand	price	user_id	user_session
0	2019-11-01 00:00:00	view	1003461	electronics	smartphone	smartphone	xiaomi	489.07	520088904	4d3b30da-a5e4-49df-b1a8-ba5943f1dd33
1	2019-11-01 00:00:00	view	5000088	appliances	sewing_machine	sewing_machine	janome	293.65	530496790	8e5f4f83-366c-4f70-860e-ca7417414283
2	2019-11-01 00:00:01	view	1306421	computers	notebook	notebook	hp	514.56	514028527	df8184cc-3694-4549-8c8c-6b5171877376
3	2019-11-01 00:00:01	view	1004775	electronics	smartphone	smartphone	xiaomi	183.27	558856683	313628f1-68b8-460d-84f6-cec7a8796ef2
4	2019-11-01 00:00:01	view	1306894	computers	notebook	notebook	hp	360.09	520772685	816a59f3-f5ae-4ccd-9b23-82aa8c23d33c

< 표1 : 다중 카테고리 온라인 상점 고객 행동 데이터셋 >

- ▶ 데이터셋 컬럼 소개

event\_time : 유저 이벤트가 발생한 시간(UTC 기준)  
event\_type : 유저 이벤트의 종류  
product\_id : 제품의 ID  
category\_code : 제품의 카테고리  
sub\_category : 제품의 서브 카테고리

product : 제품 종류  
brand : 제품의 브랜드  
price : 제품의 가격 (단위 불명)  
user\_id : 유저의 ID  
user\_session : 유저의 세션 로그

## 02. 프로젝트 수행 절차 및 방법

---

### ▶ 유저 데이터 RFM 분석

- RFM 수치를 확인하여 고객들의 전반적인 성향을 대략적으로 파악하였습니다.

### ▶ 기본적인 전처리와 정제를 거친 후, 추천 시스템 구현

- 콘텐츠 기반 추천 모델(CB)
- 행렬 분해(Matrix Factorization)를 이용한 협업 필터링(ALS)

모델 구현 주요 라이브러리



—



### 03. 프로젝트 수행 결과

▶ 구매 고객 행동데이터 RFM 수치 확인

- 최근 접속을 수치화한 recency, 접속 빈도를 나타내는 frequency, 총 구매금액을 나타내는 monetary값
- 최댓값과 최솟값의 범위 내에서 구획화하여 1~5점을 부여하였습니다.
- 값의 분포가 다양한 recency와 달리 frequency와 monetary는 대부분의 데이터가 1점대에 몰려 있습니다.
- 이어 전체 고객들의 접속 빈도수를 보면, 사분위 수의 50%에 해당하는 값이 7회로 나옵니다.

※ 전체 330,394명의 고객 중 대부분이 월 1~2회 구매에 그치는 것을 알 수 있고,  
방문 및 장바구니 로그를 모두 살펴봐도 온라인 물에 접속하는 횟수가 매우 적은 것을 알 수 있습니다.

	user_id	recency_R	frequency_F	monetary_M
0	225644257	2	1	1
1	253299396	1	1	1
2	256164170	4	1	1
3	296197073	5	1	1
4	299358698	2	1	1
...	...	...	...	...
330389	579962677	5	1	1
330390	579963523	5	1	1
330391	579964521	5	1	1
330392	579966375	5	1	1
330393	579967058	5	1	1

330394 rows × 4 columns

< 표2 : 구매 고객들의 RFM 수치 >

1	330367
2	24
3	2
5	1

Name: frequency\_F, dtype: int64

1	330310
2	69
3	12
4	2
5	1

Name: monetary\_M, dtype: int64

< 표3 : frequency, monetary의 각 value별 갯수 확인 >

count	945.000000
mean	3029.560847
std	26158.787643
min	1.000000
25%	2.000000
50%	7.000000
75%	91.000000
max	599598.000000

Name: event\_time, dtype: float64

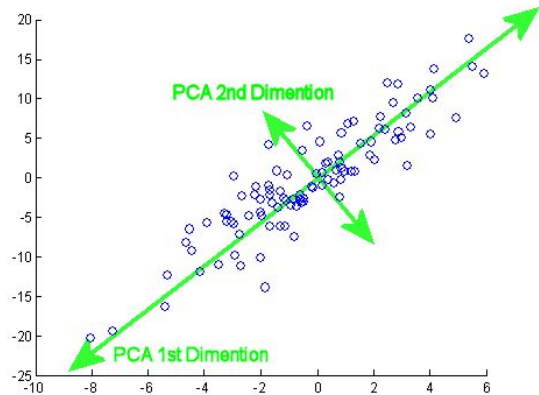
< 표4 : 전체 고객의 접속 횟수 통계치 >

### 03. 프로젝트 수행 결과

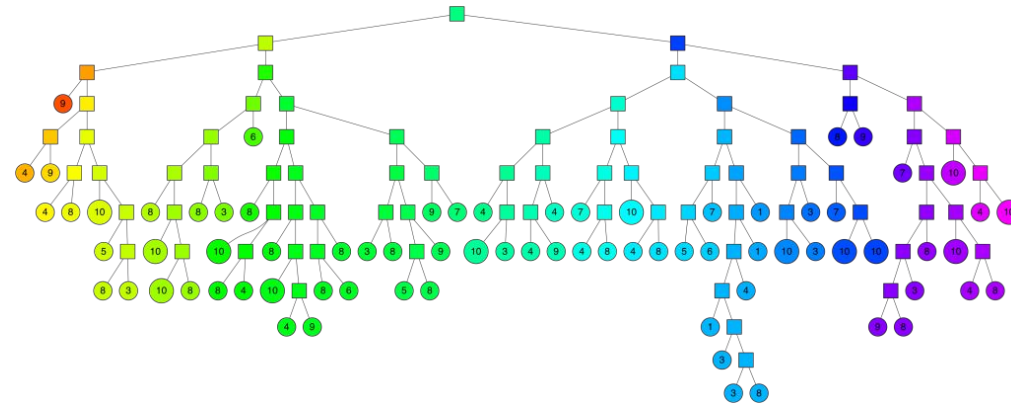
※ 접속 횟수, 구매 횟수가 적은 고객이 대다수이므로 이들에게 적절한 상품을 추천함으로써 다양한 상품을 더 확인할 수 있도록, 그리고 구매로까지 촉진하고자 합니다.

#### ▶ 콘텐츠 기반 추천 시스템

- 상품의 카테고리, 브랜드, 가격 정보를 활용하여 유사한 아이টে를 추천하는 시스템
- 카테고리 와 브랜드 데이터는 TF-IDF 활용, 가격 데이터 정규화(min-max scaling)
- 총 69,773개의 아이টে 2,159개의 특징 컬럼 -> 100개의 특징으로 PCA(메모리 절감, 속도 향상) 후 벡터화
- 벡터 유사도 검색으로 상품 간 유사도 측정 : annoy 라이브러리 활용(Approximate Nearest Neighbors)



< 그림1 : PCA >



< 그림2 : Annoy 작동 원리 >

# 03. 프로젝트 수행 결과

▶ 콘텐츠 기반 추천 시스템 구현 결과

	product_id	category_code	sub_category	product	brand	price	product_features
777	1004650	electronics	smartphone	smartphone	samsung	622.27	electronics smartphone smartphone samsung
90	1003050	electronics	smartphone	smartphone	samsung	617.52	electronics smartphone smartphone samsung
709	1004531	electronics	smartphone	smartphone	samsung	643.23	electronics smartphone smartphone samsung
1198	1005173	electronics	smartphone	smartphone	samsung	643.23	electronics smartphone smartphone samsung
1199	1005174	electronics	smartphone	smartphone	samsung	643.23	electronics smartphone smartphone samsung
1200	1005175	electronics	smartphone	smartphone	samsung	643.23	electronics smartphone smartphone samsung
1226	1005201	electronics	smartphone	smartphone	samsung	643.20	electronics smartphone smartphone samsung
263	1003710	electronics	smartphone	smartphone	samsung	643.49	electronics smartphone smartphone samsung
89	1003048	electronics	smartphone	smartphone	samsung	643.52	electronics smartphone smartphone samsung
91	1003051	electronics	smartphone	smartphone	samsung	599.66	electronics smartphone smartphone samsung
92	1003052	electronics	smartphone	smartphone	samsung	599.66	electronics smartphone smartphone samsung
398	1004016	electronics	smartphone	smartphone	samsung	589.44	electronics smartphone smartphone samsung
776	1004648	electronics	smartphone	smartphone	samsung	589.20	electronics smartphone smartphone samsung
779	1004653	electronics	smartphone	smartphone	samsung	588.92	electronics smartphone smartphone samsung
139	1003281	electronics	smartphone	smartphone	samsung	586.16	electronics smartphone smartphone samsung
205	1003507	electronics	smartphone	smartphone	samsung	586.16	electronics smartphone smartphone samsung

< 표5 : 777번 상품에 대한 15개 상품 추천 >

	product_id	category_code	sub_category	product	brand	price	product_features
36235	13903348	construction	components	faucet	lava	191.00	construction components faucet lava
36231	13903343	construction	components	faucet	lava	197.28	construction components faucet lava
36267	13903393	construction	components	faucet	elleci	176.15	construction components faucet elleci
35173	13901420	construction	components	faucet	cobra	167.58	construction components faucet cobra
36069	13903028	construction	components	faucet	damixa	163.73	construction components faucet damixa
64204	100003671	construction	components	faucet	drgans	165.00	construction components faucet drgans
64209	100003707	construction	components	faucet	drgans	165.00	construction components faucet drgans
36075	13903045	construction	components	faucet	florentina	174.01	construction components faucet florentina
35529	13902160	construction	components	faucet	florentina	165.00	construction components faucet florentina
36194	13903259	construction	components	faucet	florentina	165.00	construction components faucet florentina
36195	13903261	construction	components	faucet	florentina	165.00	construction components faucet florentina
36196	13903262	construction	components	faucet	florentina	165.00	construction components faucet florentina
64893	100005488	construction	components	faucet	jacobdelafon	181.73	construction components faucet jacobdelafon
64930	100005650	construction	components	faucet	jacobdelafon	181.73	construction components faucet jacobdelafon
64943	100005679	construction	components	faucet	jacobdelafon	181.73	construction components faucet jacobdelafon
35537	13902183	construction	components	faucet	florentina	158.82	construction components faucet florentina

< 표6 : 36235번 상품에 대한 15개 상품 추천 >

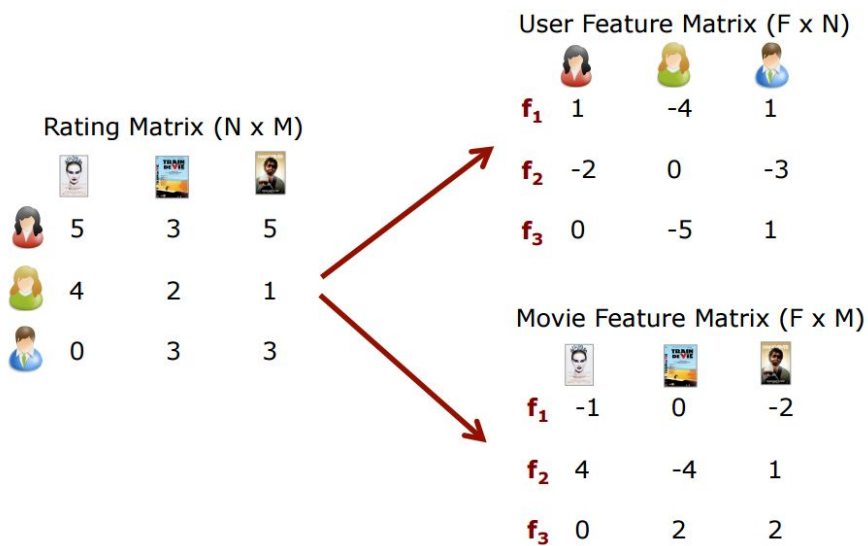
- 추천한 상품에 대한 실제 구매결과를 알 수 없어 정량적인 성능평가는 어려웠습니다.
  - 여러 상품에 대해 동작해본 결과, 기준 아이템과 유사한 아이템을 잘 찾는 것을 볼 수 있습니다.
- ※ 하지만 유사한 아이템만 결과로 나와서 추천 시스템의 우연성(Serendipity)이 없는 것으로 보입니다.  
콘텐츠 기반 추천 시스템 단독으로 고객에게 상품을 추천, 구매 욕구를 자극하기엔 조금 부족합니다.



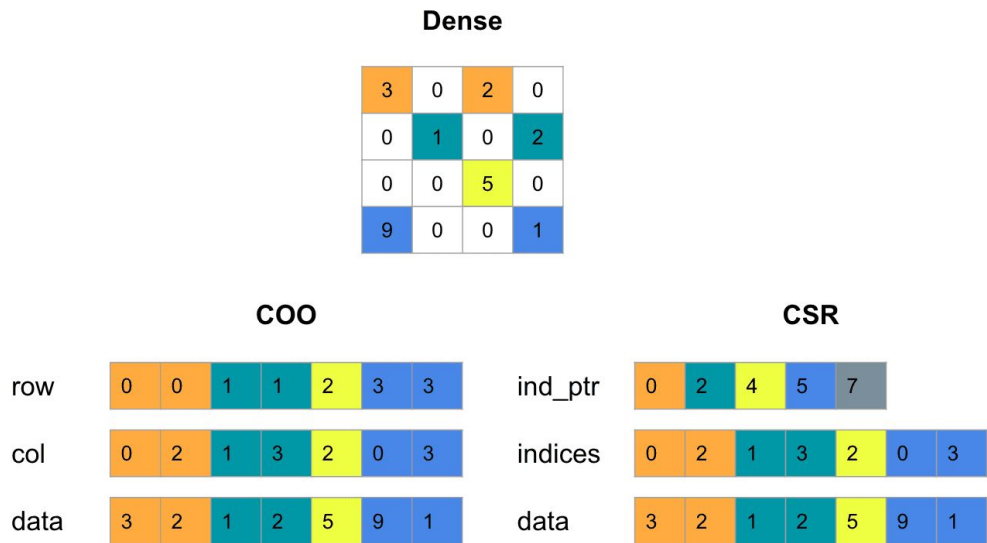
# 03. 프로젝트 수행 결과

## ▶ 행렬 분해를 이용한 협업 필터링(ALS)

- 고객들이 실제로 구매한 상품 데이터를 기반으로 행렬분해를 이용한 추천 시스템
- 구매 기록이 명시적이진 않지만 암시적으로 상품에 대한 선호도를 의미한다고 추론
- 행렬 분해를 통해 고객 - 상품 간 숨겨진 요소값을 찾아내, 상품 추천에 활용
- 총 330,394명의 고객이 구매한 20,772개의 상품 506,950건의 정보를 활용
- 20,772 x 330,394의 희소 행렬 구성 : sparsity 99.9%, 메모리 효율을 위해 CSR 행렬 구성



< 그림 3 : 행렬분해 (Matrix Factorization) >



< 그림 4 : COO, CSR Matrix >

### 03. 프로젝트 수행 결과

#### ▶ 행렬 분해를 이용한 협업 필터링(ALS) 구현

- factors = 30, regularization = 0.1, alpha = 40, iterations = 30
- 전체 matrix에서 20% 구매기록을 가린 훈련 데이터로 학습하였습니다.

#### ▶ 추천 시스템 학습 결과

- 30회 학습 시 loss = 0.000998
- 평가 : 전체 330,394명의 고객에 대한 precision@15 = 0.028



< 그림 5 : Precision@k >

```
array([ 5.6139587e-10,  6.8121175e-10,  2.0781465e-09, -1.4576713e-09,
        1.0343078e-09, -6.6087580e-11,  1.2209045e-10, -6.0763727e-10,
        9.9720388e-10, -2.7078118e-10, -4.8773985e-10, -7.0459183e-10,
       -1.2860784e-09,  5.0787607e-10,  3.1348141e-10, -1.3801738e-10,
        2.5216895e-10, -2.1469118e-09, -1.9341800e-10, -1.0804971e-09,
       -3.2890213e-10, -7.0781758e-10, -7.9733242e-10, -1.2757986e-10,
        2.9951375e-11,  1.2455992e-09,  1.9296009e-09,  6.6413897e-10,
        1.6932611e-10, -3.0788283e-10], dtype=float32)
```

< 표 7 : 330394번째 유저의 latent factors >

```
array([ 7.78323933e-11,  9.42814160e-11,  1.14283680e-10,  5.33615374e-11,
        1.06796849e-10,  8.49002188e-11,  1.13617185e-10,  2.78993495e-12,
        6.74255096e-11,  2.76100531e-11,  9.97703031e-11, -8.54378374e-12,
        1.04527768e-10,  5.89571725e-11,  6.65972277e-11,  9.92708693e-11,
        6.92488566e-11,  3.39354378e-11,  5.68227687e-11,  4.89985622e-11,
        2.14261664e-11,  1.17031634e-10,  7.06665282e-11,  2.93320437e-11,
        1.30418384e-10,  8.73323636e-12, -1.09421916e-11,  2.51326737e-12,
       -9.39945344e-12,  6.27950747e-11], dtype=float32)
```

< 표 8 : 20772번째 제품의 latent factors >

### 03. 프로젝트 수행 결과

▶ 행렬 분해를 이용한 협업 필터링(ALS) 구현 결과

	product_id	category_code	sub_category	product	brand	price	product_features
	777	1004650	electronics	smartphone	smartphone	samsung	622.27 electronics smartphone smartphone samsung
	779	1004653	electronics	smartphone	smartphone	samsung	588.92 electronics smartphone smartphone samsung
	786	1004663	electronics	smartphone	smartphone	samsung	898.32 electronics smartphone smartphone samsung
	1000	1004945	electronics	smartphone	smartphone	samsung	952.15 electronics smartphone smartphone samsung
	1168	1005142	electronics	smartphone	smartphone	apple	2004.17 electronics smartphone smartphone apple
	1198	1005173	electronics	smartphone	smartphone	samsung	643.23 electronics smartphone smartphone samsung
	1200	1005175	electronics	smartphone	smartphone	samsung	643.23 electronics smartphone smartphone samsung
	1212	1005187	electronics	smartphone	smartphone	samsung	771.94 electronics smartphone smartphone samsung
	8866	2900847	appliances	kitchen	microwave	bbk	97.79 appliances kitchen microwave bbk
	11732	4200485	appliances	environment	air_conditioner	almacom	1014.82 appliances environment air_conditioner almacom
	18824	6200327	appliances	environment	air_heater	electrolux	25.71 appliances environment air_heater electrolux
	28732	10400735	kids	toys	toys	wincars	81.86 kids toys toys wincars
	37695	16800028	furniture	kitchen	table	sv	99.87 furniture kitchen table sv
	48921	28400415	accessories	bag	bag	petek	155.86 accessories bag bag petek
	49346	28401346	accessories	bag	bag	picard	110.68 accessories bag bag picard
	58963	35200685	apparel	underwear	underwear	visavis	7.79 apparel underwear underwear visavis

< 표9 : 777번 상품에 대한 15개 상품 추천2 >

- 앞서 본 CB모델보다 조금 더 다양한 상품을 추천하는 것을 볼 수 있습니다.
- 고객의 성향이 Matrix에 반영되어 어느정도 개인화 추천도 가능한 것을 볼 수 있습니다.

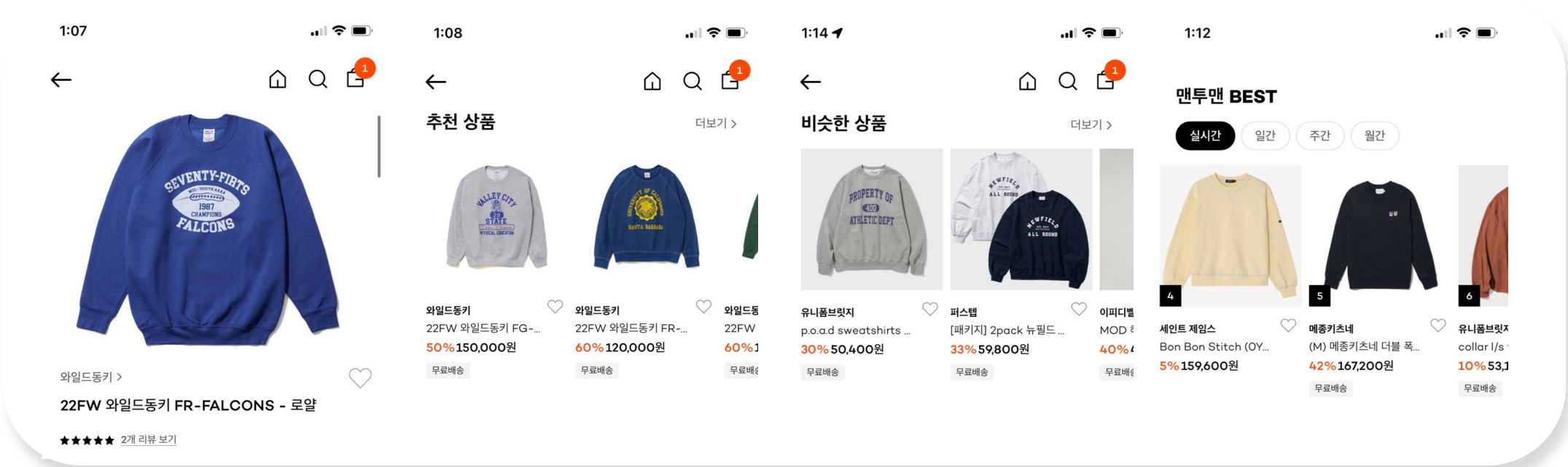
※ 정성적으로 평가했을 때, 좀 더 다양한 추천이 가능한 ALS모델이 CB모델보다 적용하기 좋을 것으로 보입니다.  
다만 유저의 구매기록이 없는 경우, CB모델로 우선 유사 아이টে을 추천하는 것도 좋을 것 같습니다.

	product_id	category_code	sub_category	product	brand	price	product_features
	51	1002629	electronics	smartphone	smartphone	apple	358.31 electronics smartphone smartphone apple
	265	1003712	electronics	smartphone	smartphone	samsung	720.22 electronics smartphone smartphone samsung
	729	1004565	electronics	smartphone	smartphone	huawei	164.95 electronics smartphone smartphone huawei
	851	1004750	electronics	smartphone	smartphone	samsung	195.06 electronics smartphone smartphone samsung
	917	1004833	electronics	smartphone	smartphone	samsung	172.23 electronics smartphone smartphone samsung
	919	1004835	electronics	smartphone	smartphone	samsung	223.95 electronics smartphone smartphone samsung
	920	1004836	electronics	smartphone	smartphone	samsung	229.90 electronics smartphone smartphone samsung
	961	1004888	electronics	smartphone	smartphone	samsung	224.46 electronics smartphone smartphone samsung
	1185	1005159	electronics	smartphone	smartphone	xiaomi	212.10 electronics smartphone smartphone xiaomi
	1194	1005168	electronics	smartphone	smartphone	samsung	213.94 electronics smartphone smartphone samsung
	1195	1005169	electronics	smartphone	smartphone	samsung	211.61 electronics smartphone smartphone samsung
	1230	1005205	electronics	smartphone	smartphone	oppo	256.74 electronics smartphone smartphone oppo
	1242	1005217	electronics	smartphone	smartphone	xiaomi	151.07 electronics smartphone smartphone xiaomi
	1262	1005238	electronics	smartphone	smartphone	oppo	308.45 electronics smartphone smartphone oppo
	16316	4804660	electronics	audio	headphone	xiaomi	23.05 electronics audio headphone xiaomi

< 표10 : ID 519261186 유저가 선호할 만한 아이টে을 추천 >

# 03. 프로젝트 수행 결과

▶ 추천 시스템 적용 예시



< 그림6 : 추천 시스템 적용 예시(모바일 쇼핑몰 29cm 화면) >

- 예시처럼 다양한 관점으로 상품을 추천하는 것이 좋아 보입니다.

※ 같은 카테고리 내에서 가장 많이 팔린 상품을 추천하는 것도 효과적인 추천방식으로 보입니다.

## 04. 자체 평가 의견

---

### ▶ 아쉬운 점

- 유저 또는 아이템의 메타데이터에 한계가 있고, 그에 비해 데이터 규모가 커서 추천 시스템 구현에 다소 제약이 있었습니다.
- 좀 더 다양한 추천모델을 구현해보지 못한 아쉬움이 있습니다. (best 상품 추천, 딥러닝 모델 등)
- 과정 중 발생하는 다양한 문제와 오류를 효율적으로 해결하지 못해 전반적으로 프로젝트 진행이 더뎠고 완성도가 떨어지지 않았나 생각합니다.

### ▶ 뜻 깊은 점

- 대용량 데이터를 효율적으로 처리할 수 있는 여러 방법들을 활용해볼 수 있었습니다.
- 기본적이고 중요한 추천 시스템에 대해 공부하고 과정 전반을 이해할 수 있었습니다.