

# 고객 로그 데이터를 활용한 추천 모델링

전현아

# 목차

- 01. 프로젝트 배경**
- 02. 프로젝트 팀 구성 및 역할**
- 03. 프로젝트 수행 절차**
- 04. 데이터 소개 및 분석**
- 05. 추천시스템 수행 결과**
- 06. 프로젝트 회고**

## 01. 프로젝트 배경

### 지금은 이커머스 저성장 시대



사진출처

코로나19 대유행으로 비대면 문화가 확산되고 디지털 전환이 가속화됨에 따라 이커머스 시장은 급격한 성장을 이루었습니다. 하지만 2017년 이후 연평균 20%대의 고성장을 기록하던 이커머스 시장이 저성장 구간에 진입하는 것으로 보입니다. 증권가는 올해 국내 온라인 쇼핑 시장의 성장률이 9~13% 수준으로 하락할 것으로 전망합니다.

따라서 이런 시대에서 살아남기 위해 우리는 이커머스 고객 행동 데이터를 기반으로 추천시스템을 구축해 매출상승을 목표로 하고 있습니다.

## 02. 프로젝트 팀 구성 및 역할

---

이름	역할	담당업무
전현아	팀원	<ul style="list-style-type: none"><li>▶ 고객 세분화</li><li>▶ 추천시스템 구축</li><li>▶ ppt 및 기술보고서 작성</li></ul>
홍영신	팀원	<ul style="list-style-type: none"><li>▶ 데이터 경량화</li><li>▶ 추천시스템 구축</li></ul>

### 03. 프로젝트 수행 절차

구분	기간	활동
사전 기획	▶ 1/31(화)	▶ 프로젝트 기획
데이터 분석 및 전처리	▶ 2/1(수) ~ 2/2(목)	▶ EDA ▶ 데이터 전처리 및 경량화
모델링	▶ 2/3(금) ~ 2/10(금)	▶ 고객 세분화 ▶ 각 고객군에 맞는 모델 구현
모듈화	▶ 2/13(월) ~ 2/14(화)	▶ 모델 서빙을 위한 모듈화
총 개발기간	▶ 1/31(화) ~ 2/15(수) (총 2주)	-

## 04. 데이터 소개 및 분석

---

### 데이터 소개

- click\_stream\_new.csv : 고객 로그 데이터
- transaction\_new.csv : 고객 구매 데이터
- customer.csv : 고객 정보
- product.csv : 제품 정보
- 출처

<https://www.kaggle.com/datasets/latifahhukma/fashion-campus>

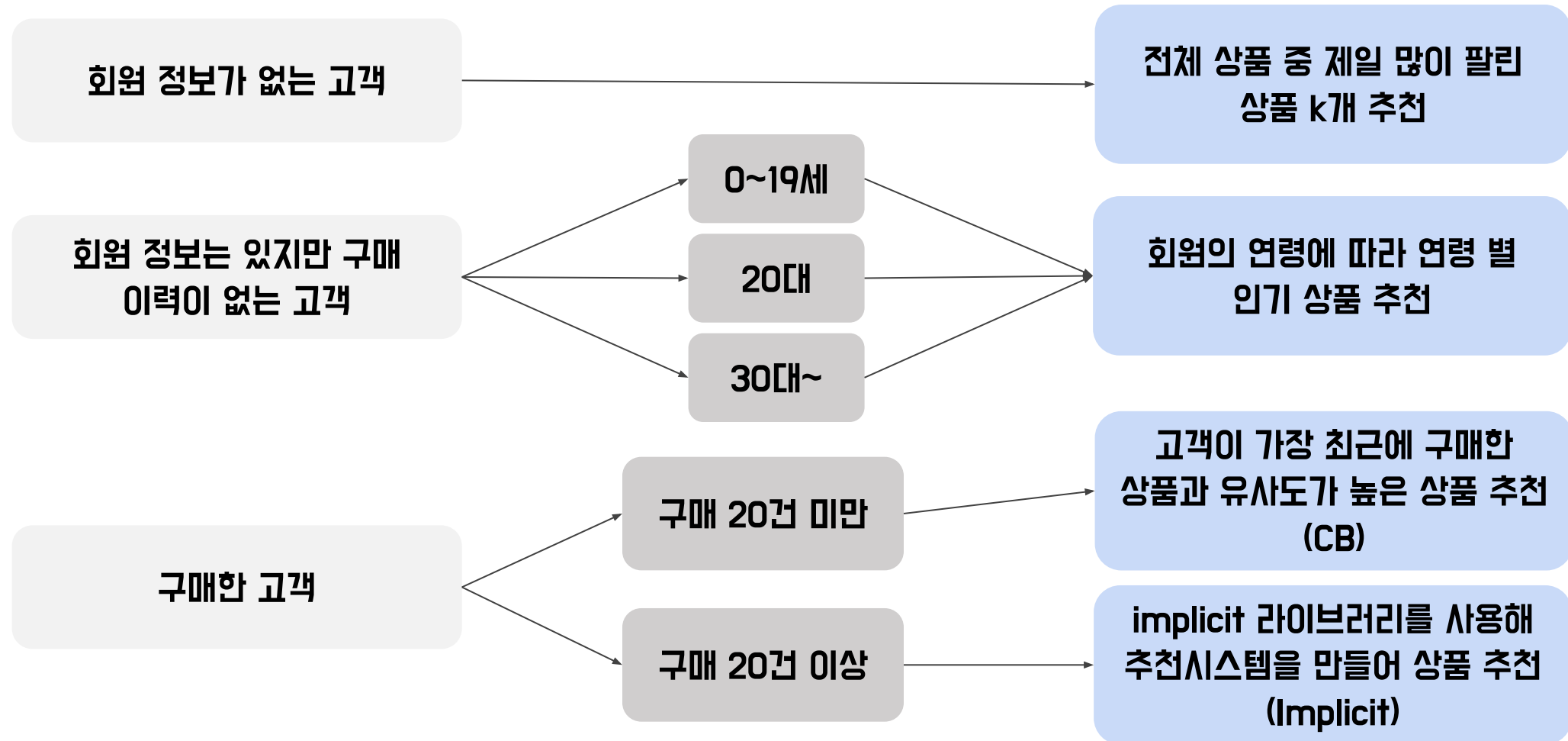
### 데이터 경량화

- object -> category로 변환
- float64, int64를 길이에 맞게 더 작은 크기로 변환
- click\_stream\_new : 1.1GB -> 467.45MB (60.2% 감소)
- transaction\_new : 153.1MB -> 110.07MB (28.1% 감소)
- customer : 11.4MB -> 5.01MB (56.3% 감소)
- product : 3.4MB -> 0.86MB (74.7% 감소)
- csv -> parquet로 저장

## 04. 데이터 소개 및 분석

### 고객 세분화

### 추천 모델



## 05. 추천시스템 수행 결과(CB)

고객이 가장 최근에 구매한 제품1개와 유사한 제품 25개 출력

### 고객이 가장 최근에 구매한 상품

product_id	info
11834	Men/ <b>Footwear/Shoes</b> /Casual Shoes/ <b>Grey</b>

product_id	info	product_id	info
15970	Men/Apparel/Topwear/Shirts/ <b>Navy Blue</b>	29928	Men/Accessories/Watches/Watches/ <b>Black</b>
39386	Men/Apparel/Bottomwear/Jeans/Blue	17429	Men/Accessories/Watches/Watches/ <b>Black</b>
21379	Men/Apparel/Bottomwear/Track Pants/ <b>Black</b>	12967	Men/ <b>Footwear/Sandal</b> /Sandals/ <b>Black</b>
53759	Men/Apparel/Topwear/Tshirts/ <b>Grey</b>	6842	Unisex/Accessories/ <b>Shoe Accessories/Shoe Acces...</b>
1855	Men/Apparel/Topwear/Tshirts/ <b>Grey</b>	13089	Men/Apparel/Topwear/Sweatshirts/ <b>Grey</b>
30805	Men/Apparel/Topwear/Shirts/Green	18461	Men/Personal Care/Fragrance/Deodorant/White
29114	Men/Accessories/ <b>Socks/Socks</b> / <b>Navy Blue</b>	9036	Men/ <b>Footwear/Shoes</b> /Formal Shoes/ <b>Black</b>
30039	Men/Accessories/Watches/Watches/ <b>Black</b>	7990	Men/Apparel/Topwear/Tshirts/ <b>Navy Blue</b>
9204	Men/ <b>Footwear/Shoes</b> /Casual Shoes/ <b>Black</b>	37812	Men/Apparel/Topwear/Shirts/ <b>Navy Blue</b>
18653	Men/ <b>Footwear/Flip Flops</b> /Flip Flops/ <b>Black</b>	4729	Boys/Apparel/Topwear/Tshirts/Green
46885	Boys/ <b>Footwear/Flip Flops</b> /Flip Flops/ <b>Navy Blue</b>	56825	Men/Apparel/Topwear/Shirts/Brown
12369	Men/Apparel/Topwear/Shirts/Purple	39988	Men/ <b>Footwear/Shoes</b> /Casual Shoes/White
		18839	Men/Accessories/Belts/Belts/ <b>Black</b>



## 05. 추천시스템 수행 결과(ALS)

---

### Implicit 라이브러리를 이용한 추천시스템

#### Implicit feedback data

---



샀거나 장바구니에 담았다면 값을 1로 둬

#### 베이스 라인

---

가장 많이 팔린 25개와 비교 (50000명)

mean\_precision@k : 0.0139

mean\_hit\_rate@k : 0.2644

## 05. 추천시스템 수행 결과(ALS)

최적의 하이퍼파라미터를 위해  
Grid, Random search 진행

factor	alpha	regularization	hit_rate	precision
180	1	32	0.282189	0.016061
160	1	30	0.280843	0.016043
160	1	32	0.280395	0.016025
200	1	32	0.279498	0.015953
180	1	30	0.280395	0.015917

최종 모델 성능

mean\_precision@k : 0.0158  
mean\_hit\_rate@k : 0.2825

## 05. 추천시스템 수행 결과(ALS)

### 고객이 구매한 상품 정보

Men/Apparel/Bottomwear/J Jeans/Blue
Women/Apparel/Topwear/Tunics/Blue
Men/Footwear/Shoes/Formal Shoes/Brown
Women/Footwear/Shoes/Heels/White
Men/Apparel/Topwear/Tshirts/White
Men/Apparel/Topwear/Tshirts/Blue
Unisex/Accessories/Bags/Duffel Bag/Red
Men/Footwear/Shoes/Sports Shoes/White
Men/Footwear/Shoes/Sports Shoes/White
Women/Apparel/Topwear/Kurtas/Purple
Men/Footwear/Shoes/Sports Shoes/White
Unisex/Accessories/Eyewear/Sunglasses/Black
Men/Apparel/Innerwear/Briefs/Blue
Men/Apparel/Topwear/Tshirts/Purple
Men/Footwear/Shoes/Sports Shoes/White
Men/Accessories/Cufflinks/Cufflinks/Steel
Women/Apparel/Topwear/Tops/White

### 고객에게 추천한 상품 정보

Men/Footwear/Shoes/Formal Shoes/Black
Women/Footwear/Shoes/Heels/Black
Men/Footwear/Shoes/Formal Shoes/Brown
Women/Accessories/Bags/Handbags/Black
Women/Footwear/Shoes/Heels/Brown
Men/Footwear/Shoes/Formal Shoes/Black
Women/Accessories/Bags/Handbags/Brown
Women/Footwear/Shoes/Heels/Black
Women/Footwear/Shoes/Flats/Black
Men/Footwear/Shoes/Casual Shoes/Khaki

고객이 구매한 것은 주로 **신발**임.  
그래서 주로 신발을 추천해주는 것을 알 수 있음.

## 06. 프로젝트 회고

---

- fashion campus 데이터 중 click\_stream\_new에서 어떤 상품을 클릭했는지 봤는지 나와있지 않았음. 그래서 상품아이디가 존재하는 add\_to\_cart만 사용했음. 고객의 관심상품을 자세히 알 수 없어 아쉬웠음.
- 상품 데이터가 너무 방대해 hit rate, precision@k가 많이 낮게 나왔음.
- LightFM 모델도 사용했지만 똑같은 상품을 추천해 역확률가중치를 이용했지만 똑같은 결과가 나왔음.
- 추천시스템이 많이 재밌어져 더 많은 모델에 대해 공부해 볼 예정임.
- 이번 기회에 이커머스 관련 도메인과 추천시스템에 대해 많은 것을 알고 배울 수 있어 뿌듯함.
- 팀프로젝트를 진행하면서 부족했던 점들을 서로 보완해주고 혼자 한 것보다 만족할 수 있는 프로젝트를 만들 수 있었음.

## 깃허브 링크

---

[깃허브 링크](#)

[노션 기술 보고서 링크](#)

## 참고자료

---

[2022 엔테믹 이커머스 저성장 시대에서 살아남기 - D2C\(Direct To Consumer\) 비즈니스를 시작해야 하는 이유](#)

[pyconkr2019\\_똥똥하고 굶픈 판다를 위한 효과적인 다이어트 전략\\_20190817](#)

**이상으로 발표 마치겠습니다.**

**들어주셔서 감사합니다!**