

2024 REPORT

Passion makes Fashion(PMF)

FashionFinder

TABLE OF CONTENTS

목차 소개

01 | 주제

02 | 팀원 및 역할

03 | 사용 기술 및 개발환경

04 | 협업

05 | 데이터 파이프라인

06 ~ 11 | 진행과정

- 딥러닝을 활용한 옷 검색 웹사이트 제작
- Fasi DL 이라는 프로젝트를 참고
- <https://6mini.github.io/project/2021/11/13/fasi/>

02 | 팀원 및 역할

01 | 김정환(팀장)

- 이미지 모델링, DB

02 | 김민수

- 텍스트 모델링

03 | 최성호

- 이미지 모델링

04 | 고서연

- 웹 개발

- 공동 작업: 데이터 수집, 전처리

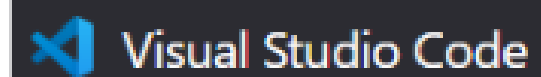
03

사용 기술 및 개발 환경

데이터 관리



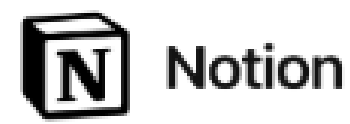
개발 도구



개발 언어 및 프레임 워크



협업 툴



04 | 협업

▼ 모델링 코드 5

Aa	👤 생성자	🕒 생성 일시	👤 최종 편집자	+	...
이미지 전처리(남자) 💬 1	👤 jjhwk	2024년 5월 21일 오전 11:27	👤 jjhwk		
resnet50 이미지모델링 💬 1	👤 jjhwk	2024년 5월 21일 오전 11:29	👤 jjhwk		
VGG19 model test 코드(이미지) 💬 3	👤 성호 최	2024년 5월 23일 오전 11:46	👤 jjhwk		
이미지 스타일 분류 + 이미지 유사도 파악 💬 1	👤 jjhwk	2024년 5월 24일 오전 11:38	👤 jjhwk		
텍스트 임베딩 코드 💬 1	👤 민수 김	2024년 5월 30일 오후 2:30	👤 민수 김		
+ 새로 만들기					

▼ 모델컷 DATA 2

Aa	👤 생성자	🕒 생성 일시	👤 최종 편집자	+	...
📄 men_style_all.csv 💬 1	👤 jjhwk	2024년 5월 21일 오전 11:31	👤 jjhwk		
women_style_all.csv 💬 2	👤 성호 최	2024년 5월 22일 오전 10:16	👤 jjhwk		
+ 새로 만들기					

▼ 크롤링 코드 2

Aa	👤 생성자	🕒 생성 일시	👤 최종 편집자	+	...
옷 상세정보 크롤링 코드(남자) 💬 2	👤 민수 김	2024년 5월 22일 오전 9:13	👤 jjhwk		
옷 상세정보 크롤링 코드(여자) 💬 1	👤 jjhwk	2024년 5월 28일 오전 9:26	👤 jjhwk		

- 노션을 통해 데이터 및 코드 공유
- 공유한 코드와 상세 내용 바탕으로 피드백

05

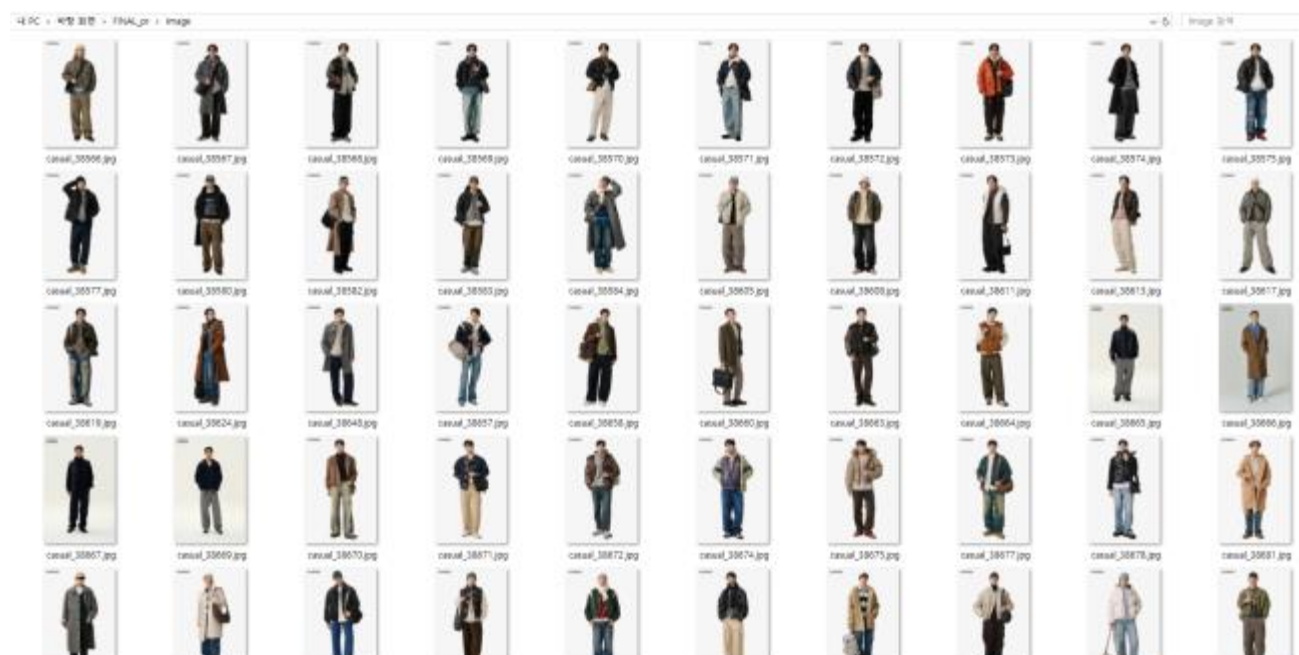
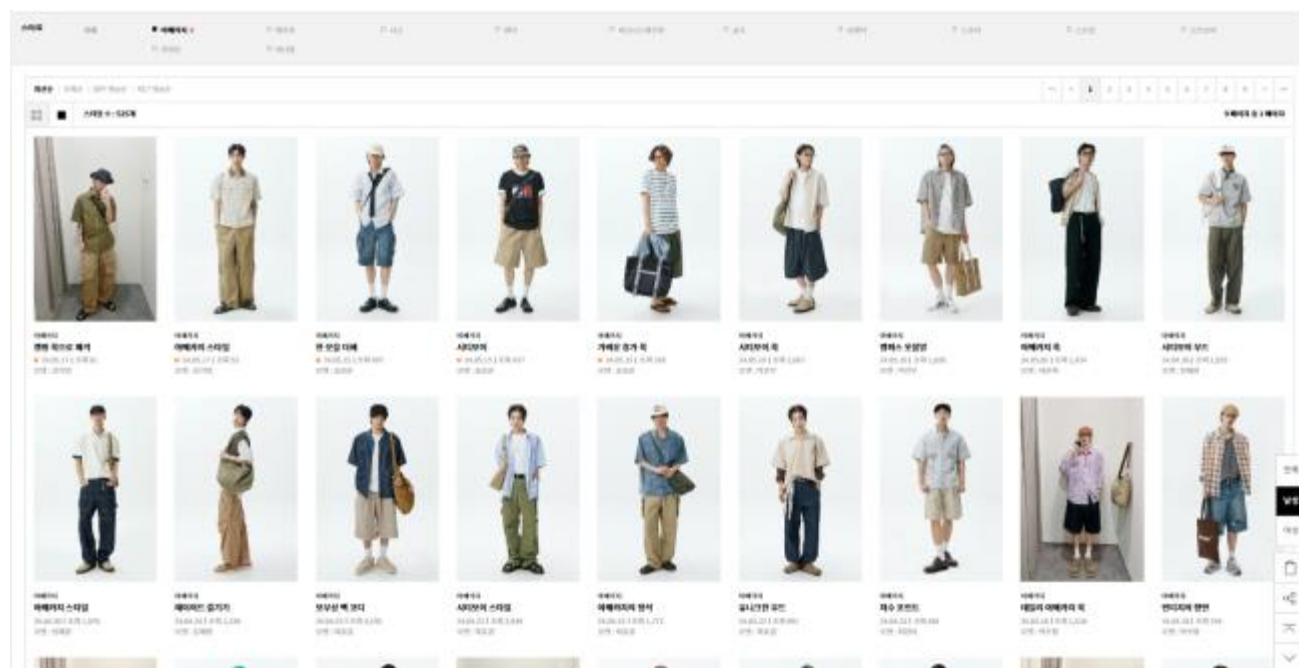
데이터 파이프라인



- 무신사 데이터 수집
- VGG19를 사용하여 이미지 분류
- ResNet-50 을 사용하여 이미지 분류 및 유사 이미지 검색
- SBERT를 사용하여 유사 상품 텍스트 검색
- Django로 웹 구현

06

데이터 수집 (스타일 페이지)



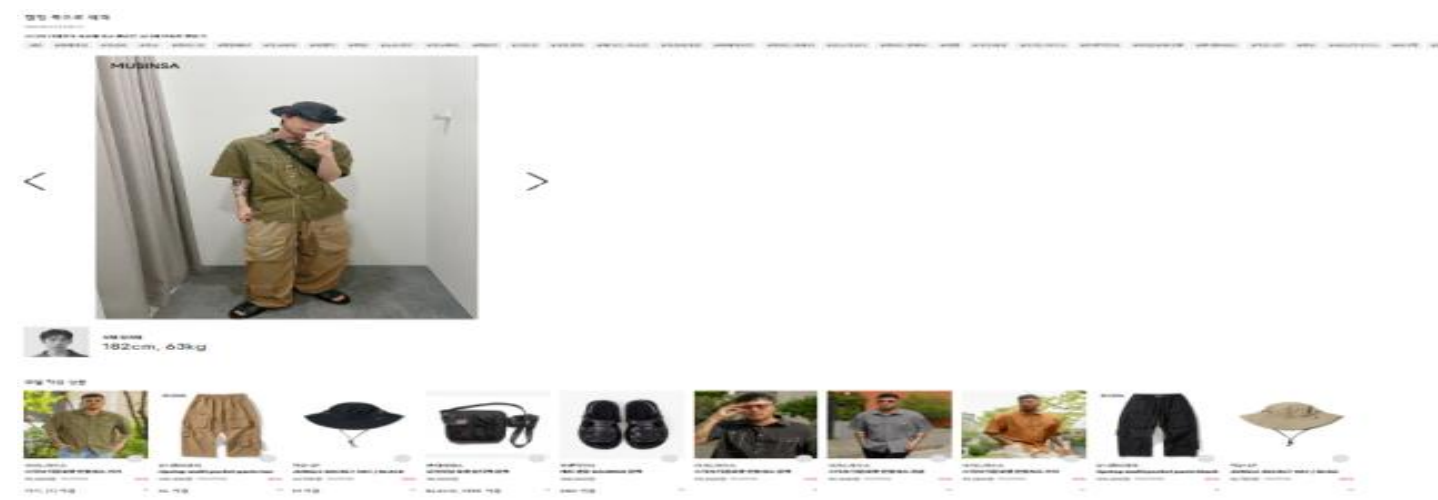
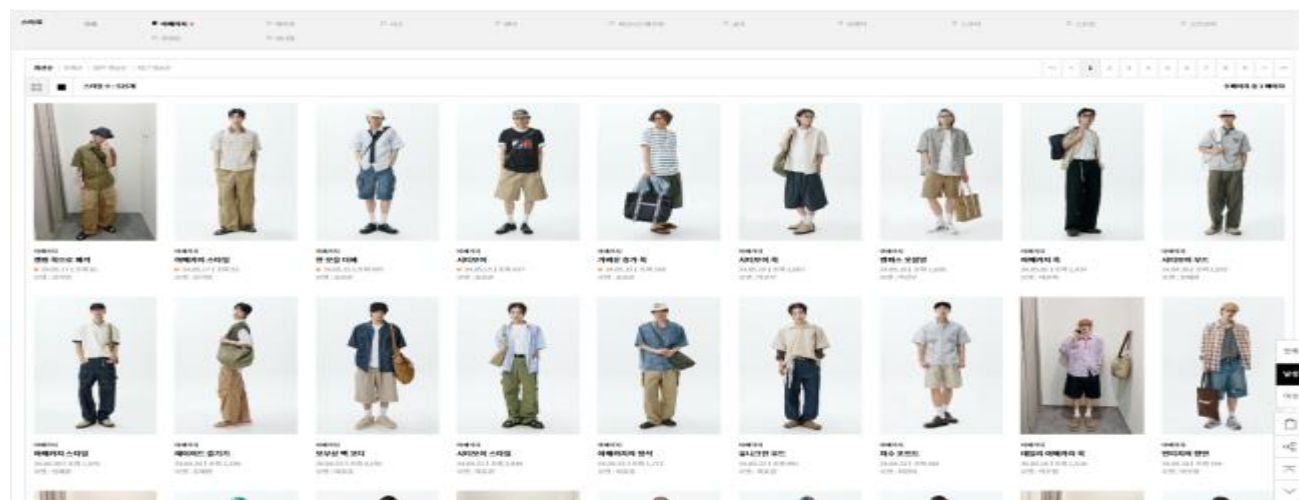
- 수집한 데이터 : 모델 이미지의 url, 스타일 종류(아메카지, 케주얼, 스트릿, ...), 모델이 입고있는 옷들로 이동 할 수 있는 id 값
- 수집 방식 : 동적 크롤링

```
all_info_df.iloc[210:220]
```

	0	1	2	3	4	5	6	7
210	아메카지	https://image.msscdn.net/images/style/detail/3...	3593776	3486091	3566041	3583355	3392943	3574622
211	아메카지	https://image.msscdn.net/images/style/detail/3...	3565855	3609284	3589056	2891608	3591782	None
212	아메카지	https://image.msscdn.net/images/style/detail/3...	2817177	3593787	3609934	3301482	2797035	2784069
213	아메카지	https://image.msscdn.net/images/style/detail/3...	3606251	3548530	3519999	3612843	3583738	None
214	아메카지	https://image.msscdn.net/images/style/detail/3...	3576285	2793468	3566157	2646647	3549493	None
215	아메카지	https://image.msscdn.net/images/style/detail/3...	2817174	3545827	3519996	3471522	3552195	3549494
216	아메카지	https://image.msscdn.net/images/style/detail/3...	2874984	3421979	3549493	3008619	2749385	3513362
217	아메카지	https://image.msscdn.net/images/style/detail/3...	3491003	3491497	1183598	3594230	None	None
218	아메카지	https://image.msscdn.net/images/style/detail/3...	3589070	3491527	3474709	3464618	3574501	3567425
219	아메카지	https://image.msscdn.net/images/style/detail/3...	3540275	3498129	3540208	3529940	3126200	3563109

06

데이터 수집 (스타일 페이지)



	abc style	abc style_id	abc url	abc clothe1	abc clothe2	abc clothe3	abc clothe4	abc clothe5	abc clothe6
1	아메카지	42201	https://image.msscdn.net/images/style/detail/42201/detail_42201_6645da0b2ea29_500.jpg	4093555	3840627	3870908	1936408	1903229	4093561
2	아메카지	42189	https://image.msscdn.net/images/style/detail/42189/detail_42189_6645ca84dbfaf_500.jpg	4034679	4071498	4071363	2946936	1903229	4034681
3	아메카지	42037	https://image.msscdn.net/images/style/detail/42037/detail_42037_6641d1e70b235_500.jpg	3758318	3942855	1022014	3862384	3582752	3710432
4	아메카지	42033	https://image.msscdn.net/images/style/detail/42033/detail_42033_6641ccb4c6a04_500.jpg	3935326	3997269	3818756	4027928	4032620	3289649
5	아메카지	42027	https://image.msscdn.net/images/style/detail/42027/detail_42027_6641bb3b56b4b_500.jpg	3941388	3868407	4033727	1960610	3984988	3971161
6	아메카지	42023	https://image.msscdn.net/images/style/detail/42023/detail_42023_663ce25262fff_500.jpg	4022918	3899325	4068223	3432919	3767266	3974502
7	아메카지	42021	https://image.msscdn.net/images/style/detail/42021/detail_42021_663ce0a6b65ff_500.jpg	3842914	3575320	3942846	4057058	3476935	4076398
8	아메카지	41850	https://image.msscdn.net/images/style/detail/41850/detail_41850_6632e4c79da66_500.jpg	3942553	4008283	3971652	2731146	3844510	3904824
9	아메카지	41728	https://image.msscdn.net/images/style/detail/41728/detail_41728_662fa07bc54bd_500.jpg	3926660	3464739	3900901	3710432	3962432	4005637
10	아메카지	41726	https://image.msscdn.net/images/style/detail/41726/detail_41726_662f9dd959c1d_500.jpg	4008435	4035116	3982989	3647928	3990719	3962704

- 수집한 데이터 : 스타일 종류(아메카지, 케주얼, 스트리트, ...), 모델 이미지의 url, 모델이 입고있는 옷들로 이동 할 수 있는 id 값(6개)
- 수집 방식 : 동적 크롤링

06

데이터 수집(옷 상세 페이지)

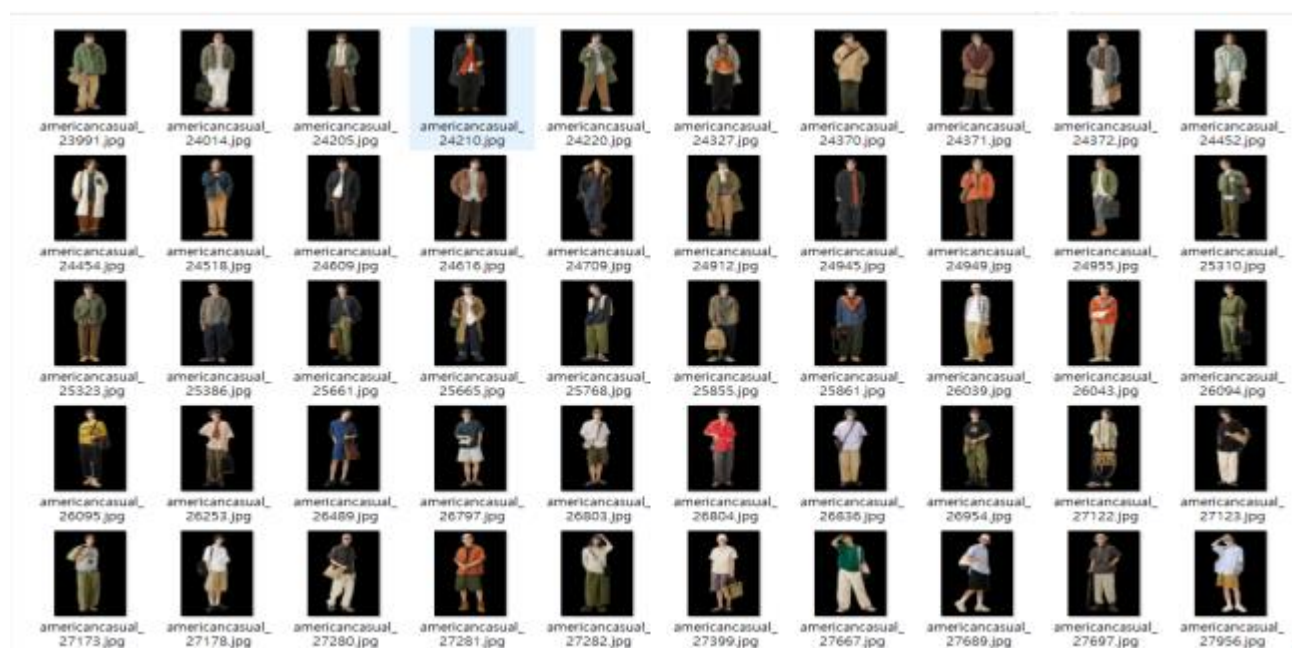
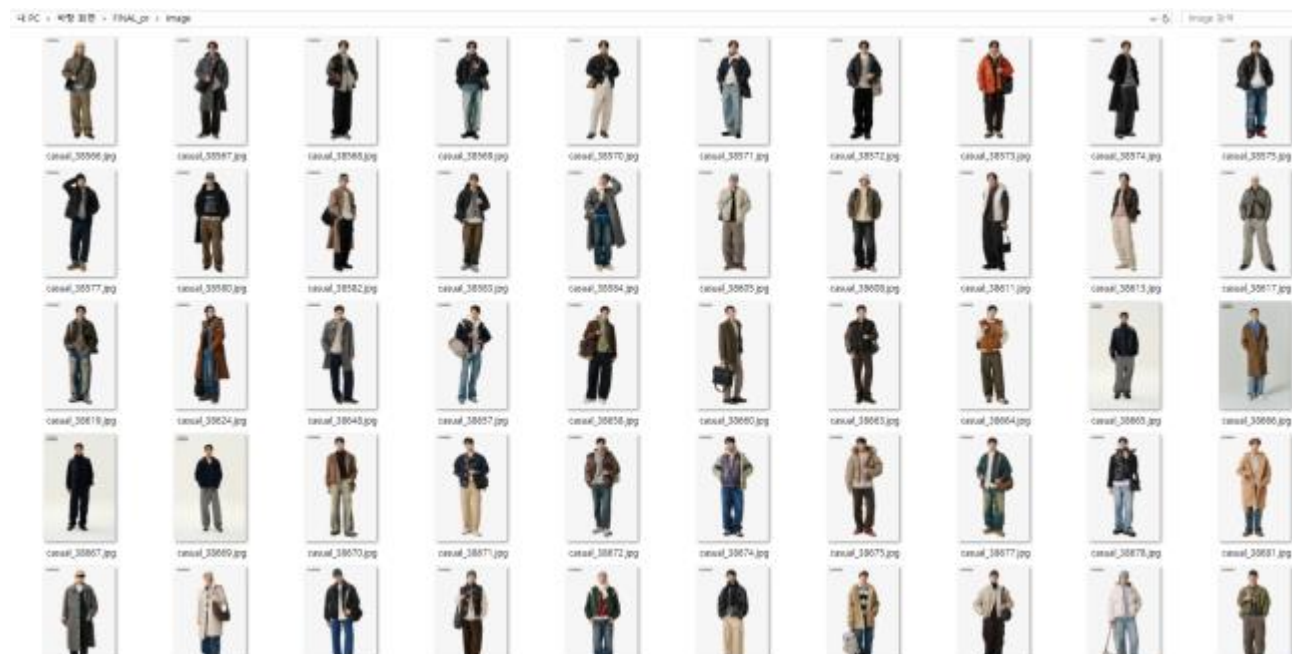


	ABC id	ABC main	ABC sub	ABC title	ABC img_url_id
1	1798977	상의	셔츠/블라우스	타탄 포켓 체크 셔츠 오버핏 블랙	20210217/1798977/1798977_16903425941487_500.jpg
2	3918784	상의	셔츠/블라우스	SITP5143 오버핏 노블 셔츠_Light pink	20240304/3918784/3918784_17095181545975_500.jpg
3	3952906	아우터	블루종/MA-1	솔리드 셔링 블루종 자켓_Cream	20240314/3952906/3952906_17104682161018_500.jpg
4	1840202	아우터	후드 집업	우먼즈 크롭 후디드 스웨트 집업 [오트밀]	20210311/1840202/1840202_17104036824120_500.jpg
5	3869943	아우터	기타 아우터	텍스처 셔링 자켓, 더스티 베이지	20240214/3869943/3869943_17095278552327_500.jpg
6	3877990	상의	맨투맨/스웨트셔츠	(WOMEN) TWENTYONE Rugby SweatShirt_Navy	20240216/3877990/3877990_17080805478373_500.jpg
7	3830641	아우터	트레이닝 재킷	L9 TRACK ZIP-UP(CHARCOAL)	20240129/3830641/3830641_17065275083306_500.jpg
8	3419368	아우터	사파리/헌팅 재킷	코튼 자켓(워시드)-베이지	20230721/3419368/3419368_16920072650283_500.jpg
9	3439979	아우터	스타디움 재킷	코듀로이 래글런 자켓_네이비_OL240	20230801/3439979/3439979_16928703467932_500.jpg
10	3634845	아우터	겨울 싱글 코트	울 더플 숏 코트_BEIGE	20231017/3634845/3634845_17120443250653_500.jpg

- 수집한 데이터 : 옷 대/소 분류, 옷 이름, 옷 이미지 url
- 수집 방식 : 동적 크롤링

07

이미지 데이터 전처리



- 사용한 전처리: 배경 검정색 변환, 이미지 증강(좌우 반전, 이미지 좌우로 이동)
- 검정색으로 변환한 이유 :
 1. 모델을 학습에 있어 배경이 일정하지 않아 이미지 분류에 악영향을 끼친다고 판단하였기 때문
 2. 흰색 배경을 선택했을 때 흰색의 옷을 잘 판단하지 못하였지만 검정색 배경을 선택했을 때는 검정색 옷을 대부분 잘 판단하였기 때문

07

텍스트 데이터 전처리

```
synonyms = {
    "팬츠" : ["팬츠", "pants", "pant"],
    "블랙" : ["블랙", "black"],
    "셔츠" : ["셔츠", "shirts"],
    "데님" : ["데님", "denim", "jeans", "jean"],
    "니트" : ["니트", "knit"],
    "와이드" : ["와이드", "wide"],
    "티셔츠" : ["티셔츠", "shirt", "tee", "tshirt", "tshirts"],
    "그레이" : ["그레이", "grey", "gray"],
    "자켓" : ["자켓", "jacket", "재킷", "점퍼"],
    "블루" : ["블루", "blue"],
}
```

- 전처리 기법 : SBERT를 사용해 텍스트 임베딩, 20번 이상 나온 단어 추출
- SBERT 사용 이유 : 텍스트 데이터인 옷 이름은 하나의 단어가 아닌 여러 단어로 이루어진 문장의 형태이기 때문에 문장 임베딩을 생성하는데 특화된 SBERT 모델을 사용 (Word2Vec(단어), FastText(단어))

스티치 더블포켓 반팔셔츠 카키

나일론 스냅 퍼티그 워크 하프 셔츠 자켓
[BEIGE]_SESH011BEIGE

릴렉스드 워시드 포플린 반소매 셔츠 [스카이 블루]

스케이터 플래그 티셔츠
(T32E1TTO05TMT2BDS)

데님 로코 포켓 반팔 셔츠 인디고 라이트



스티치 더블 포켓 반 팔 셔츠 카키

나일론 스냅 퍼티 그 워크 하프 셔츠 자켓
beige_seshbeige릴렉스드 워 시드 포플린 반 소매 셔츠 스
카이 블루

스케 이 터 플래그 티셔츠 tettotmtbds

데님 로코 포켓 반 팔 셔츠 인디고 라이트



```
tensor([[ 0.1587, -0.2798, -0.0086, ..., 0.0610, 0.0627, 0.0524],
        [-0.0894, -0.0290, -0.0062, ..., -0.0180, 0.0286, -0.0328],
        [ 0.0363, 0.1306, -0.0069, ..., -0.0282, 0.1106, 0.0300],
        ...,
        [-0.0003, 0.0607, -0.0091, ..., 0.0473, 0.1791, 0.0143],
        [-0.0065, 0.1079, -0.0018, ..., 0.0732, -0.0735, -0.0243],
        [-0.1165, 0.0406, -0.0063, ..., -0.0564, -0.1246, -0.0285]])
```

08

DB 구현

men_style_all | Enter a SQL expression to filter results (use Ctrl+Space)

	style	style_id	url	clothe1	clothe2	clothe3	clothe4	clothe5	clothe6
1	아메카지	42201	https://im	4093555	3840627	3870908	1936408	1903229	4093561
2	아메카지	42189	https://im	4034679	4071498	4071363	2946936	1903229	4034681
3	아메카지	42037	https://im	3758318	3942855	1022014	3862384	3582752	3710432
4	아메카지	42033	https://im	3935326	3997269	3818756	4027928	4032620	3289649
5	아메카지	42027	https://im	3941388	3868407	4033727	1960610	3984988	3971161
6	아메카지	42023	https://im	4022918	3899325	4068223	3432919	3767266	3974502
7	아메카지	42021	https://im	3842914	3575320	3942846	4057058	3476935	4076398
8	아메카지	41850	https://im	3942553	4008283	3971652	2731146	3844510	3904824

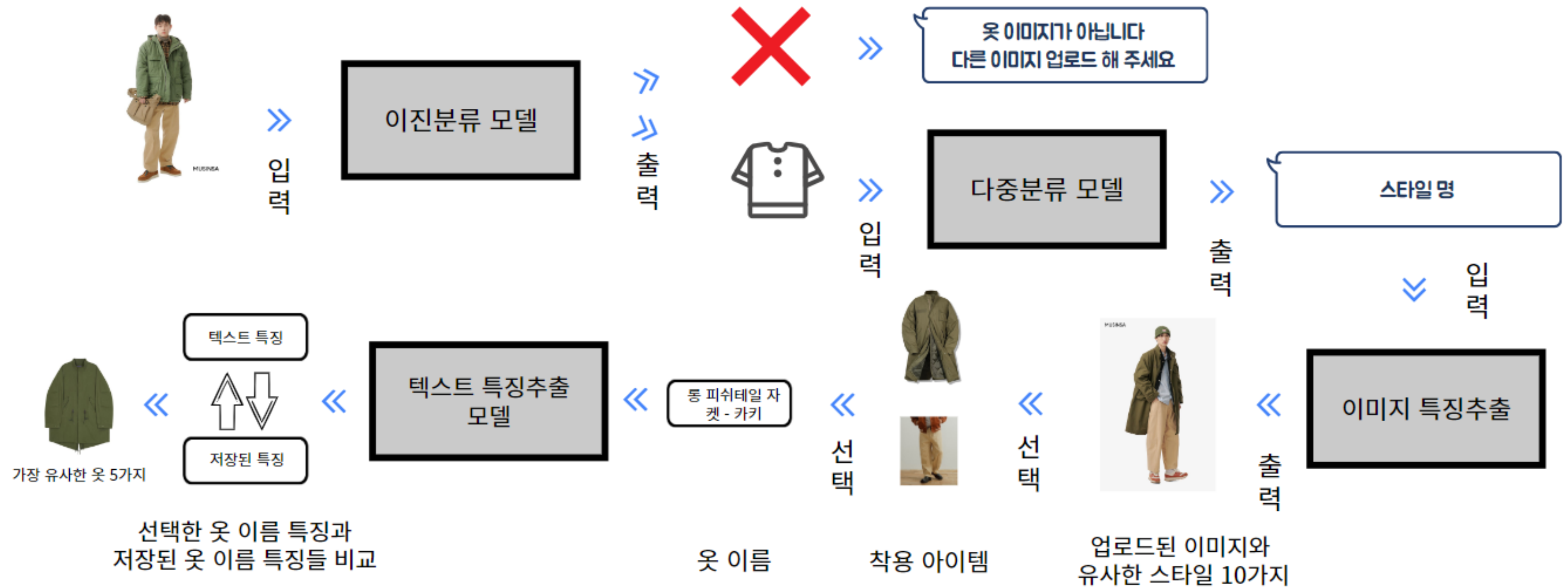
men_clothes_info | Enter a SQL expression to filter results (use Ctrl+Space)

	id	main	sub	title	img_url_id
1	409355	상의	셔츠/블라우 스티치 더블포켓 반팔셔츠 카키		20240429/4093555/4093555_1
2	403467	상의	셔츠/블라우 나일론 스냅 퍼티그 워크 하프 셔츠 자켓		20240408/4034679/4034679_1
3	375831	상의	셔츠/블라우 릴렉스드 워시드 포플린 반소매 셔츠 [20231215/3758318/3758318_1
4	393532	상의	반소매 티셔. 스케이터 플래그 티셔츠 (T32E1TTO05T		20240308/3935326/3935326_1
5	394138	상의	셔츠/블라우 데님 로코 포켓 반팔 셔츠 인디고 라이!		20240312/3941388/3941388_1
6	402291	상의	셔츠/블라우 리프 집업 셔츠 샌드 베이지		20240404/4022918/4022918_1
7	384291	상의	셔츠/블라우 집시 스트링 하프 셔츠 체크블랙		20240202/3842914/3842914_1
8	394255	상의	셔츠/블라우 트위드 반팔 셔츠 자켓_BEIGE		20240312/3942553/3942553_1

- style_all 테이블의 clothe 컬럼들을 clothes_info에서 key 값으로 사용

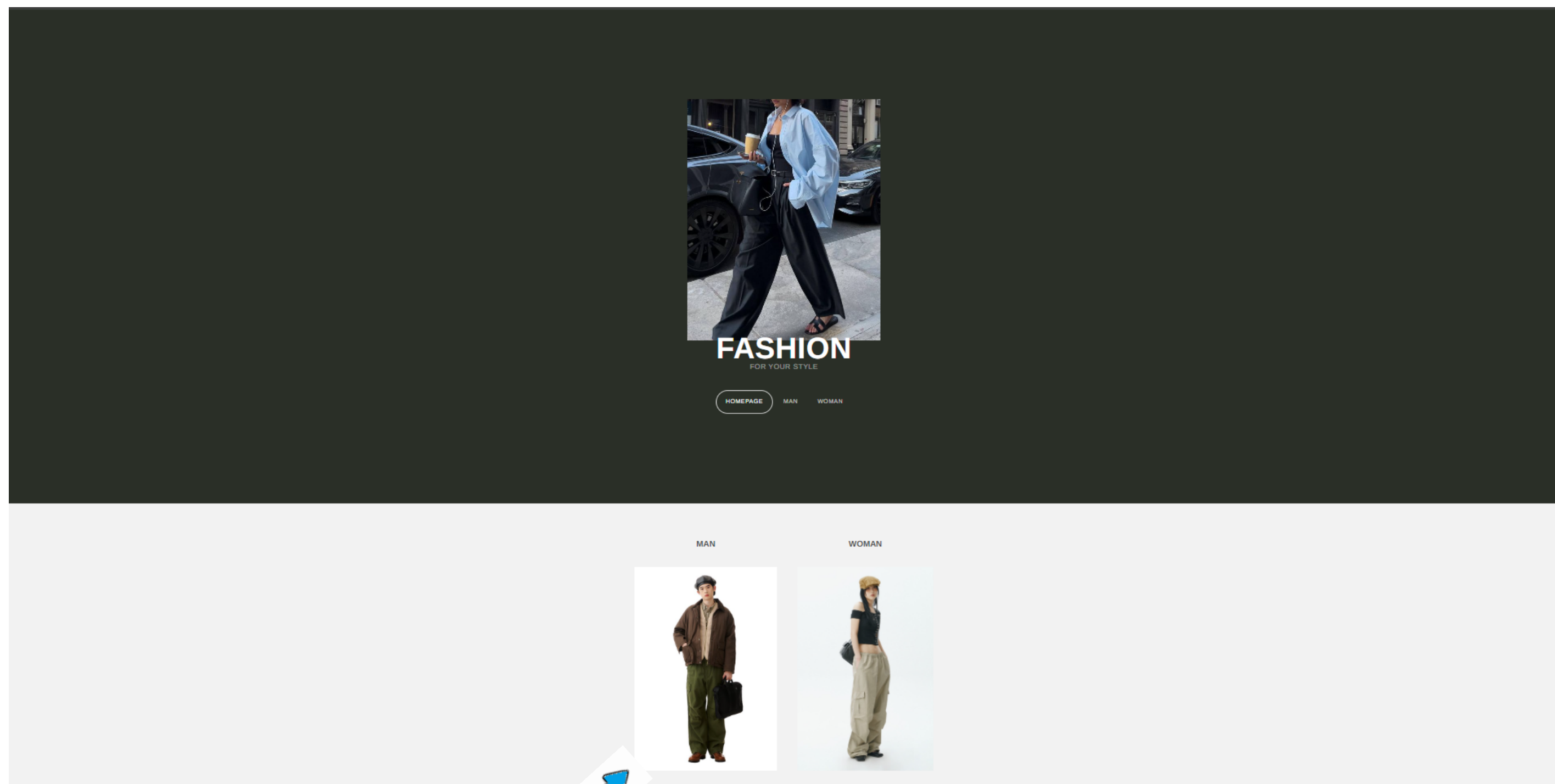
09

Web Flow Chart

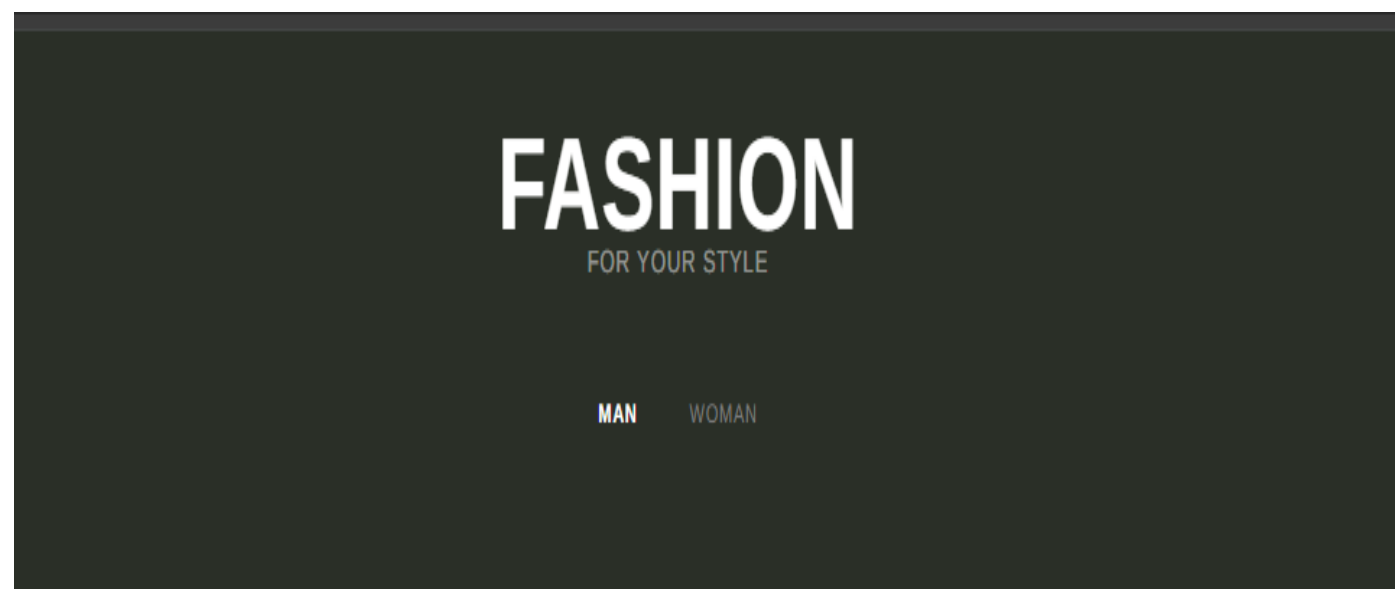


10

웹 메인 페이지



클릭 후 남자페이지로 이동



MAN UPLOAD IMAGE

이미지 선택

Upload

PREDICTION:

Uploaded Image

SIMILAR IMAGES:

PMF는 실수를 할 수 있습니다.

새로고침

이미지를 업로드하면 다음과 같이 데이터셋에 저장된 아메리칸 캐주얼 중 가장 유사한 10가지를 보여 줍니다.



업로드 후 결과 페이지

MAN UPLOAD IMAGE

이미지 선택

Upload

PREDICTION: AMERICAN



SIMILAR IMAGES:



PMF는 실수를 할 수 있습니다.

새로고침

FASHION
FOR YOUR STYLE

Clothes Worn



원하는 옷 선택

FASHION
FOR YOUR STYLE

등 피쉬테일 자켓 - 카키



원하는 옷 구매페이지로 이동

How about these clothes?



65/35 마운틴 자켓 메크루



현담자켓 VER. 2



M-65 피쉬테일 파카 카키



CHASER JACKET BARK



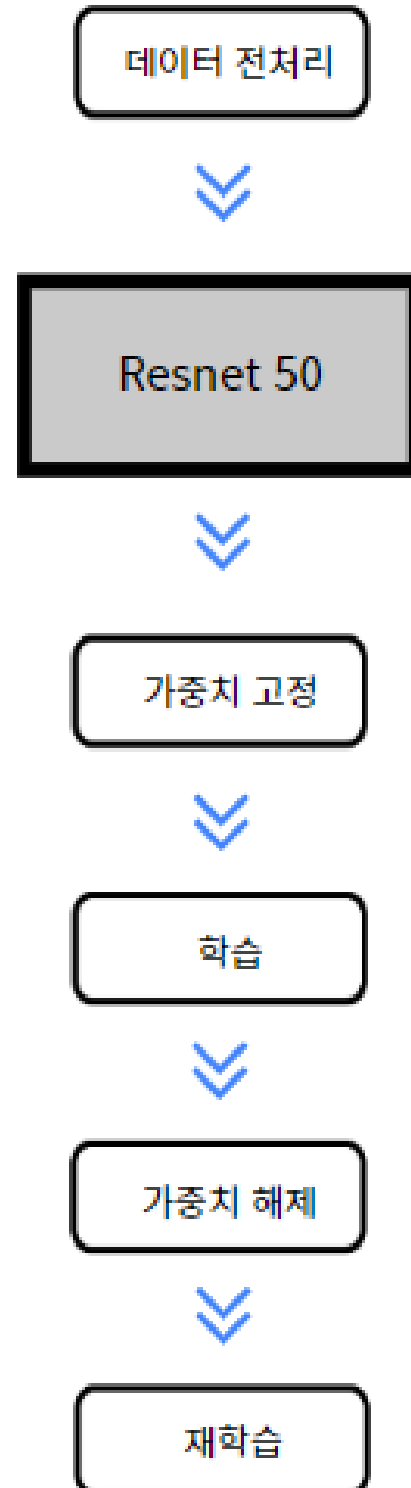
65/35 마운틴 자켓 네이비

이미지 모델(이미지 이진 분류)



- 스타일 예측에 부적합한 이미지를 걸러내기 위한 모델
- 사용 모델 : VGG19
- 이유 :
이미지가 옷인지 아닌지 구별하는 데 있어서 비교적 많은 특징 추출이 필요 없었기 때문에 VGG19의 단순한 구조가 더 효과적일 수 있다고 판단

이미지 모델링(이미지 다중 분류)



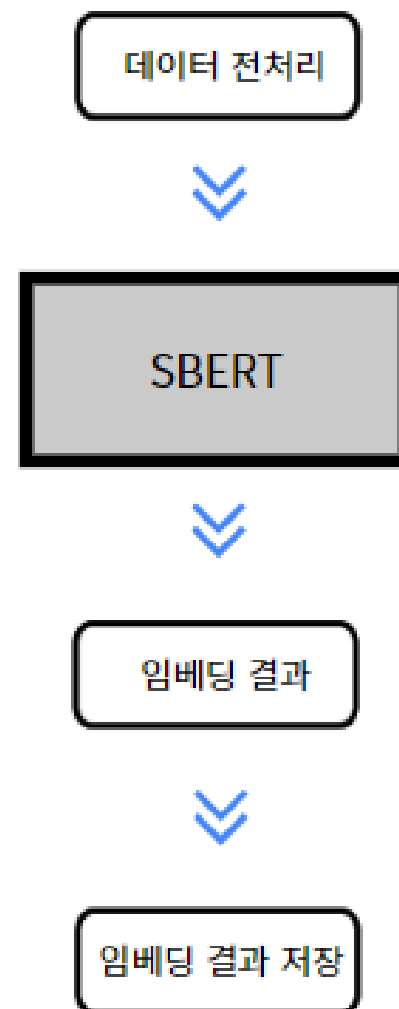
- 스타일 예측을 위한 모델
- Resnet-50 모델 사용
- 이유 :
 - 이미지가 어떤 스타일인지 구별하는데 있어서 많은 특징 추출이 필요했기 때문에
 - 비교적 복잡한 Resnet 50이 더 효과적일 수 있다고 판단
 - 또한 FASI DL 이라는 프로젝트에서도 Resnet 50을 사용하여 보다 좋은 결과를 얻었기에 사용

이미지 특징 추출(이미지 유사도 분석)



- 유사한 스타일을 찾기 위한 모델
- 사용 모델 : Resnet-50
- 사용 이유 :
 - 비슷한 이미지 찾는데 비교적 많은 특징 추출이 필요했기에 비교적 복잡한 Resnet 50이 더 효과적일 수 있다고 판단
 - 또한 코사인 유사도를 사용 했을 때 가장 속도가 빨랐기에 코사인 유사도를 사용

텍스트 모델(텍스트 임베딩)



- 텍스트 임베딩을 위한 모델
- 사용 모델 : SBERT
- 사용 이유 :
텍스트 데이터인 옷 이름은 하나의 단어가 아닌 여러 단어로 이루어진 문장의 형태이기에 문장 임베딩을 생성하는데 특화된 SBERT 모델을 사용

텍스트 특징 추출 (텍스트 유사도 분석)



- 유사한 텍스트를 찾기 위한 모델
- 코사인 유사도 사용
- 사용 이유 :

SBERT모델이 사용하는 차원은 상대적으로 고차원이고
코사인 유사도는 고차원 벡터를 다루는데 유리하므로
코사인 유사도를 사용

또한 실제로도 성능이 가장 우수함

- 보다 자세한 내용은 아래 주소에서 확인하실 수 있습니다
- <https://github.com/jjhkw>

감사합니다

Q&A