

Final Project

날씨 기반 아이 옷 추천 서비스

Machine Learning을 활용한
데이터 분석 과정

안녕
내 이름은
작은 김은*
큰 김은*
송주*
유건*
임은*
한진*

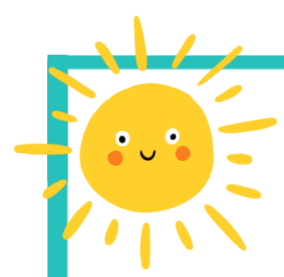


이 순서로 발표합니다

Weather Me

1. 주제 선정 배경
2. 데이터 정의
3. 데이터 처리 방안
4. 분석 결과
5. 기대효과
6. 한계점
7. Q&A





1. 주제 선정 배경

01 주제 선정 배경

Overall

Existing Applications

Necessity

Weather me

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대효과

06 한계점

07 Q & A

Q. 아이 옷을 입히실 때 무엇을 기준으로 입히시나요?

날씨죠. 아침이나 전날 저녁, 날씨를 확인하고 아이 옷을 입혀요. 예쁜 디자인의 옷은 그 다음인것 같아요. **아이**의 몸은 아직 약해서 **쉽게 감기에 걸리곤** 하거든요. 특히 환절기예요. 요즘에는 미세먼지 문제도 심각해서 **매일 날씨 체크를 하고**, 미세먼지가 심한 날에는 항상 마스크를 챙겨줘요.



정X린 (27세)
/28개월 아이 엄마

Weather Me

01 주제 선정 배경

Overall

Existing Applications

Necessity

Weather me

02 데이터 정의

03 데이터 처리 방안

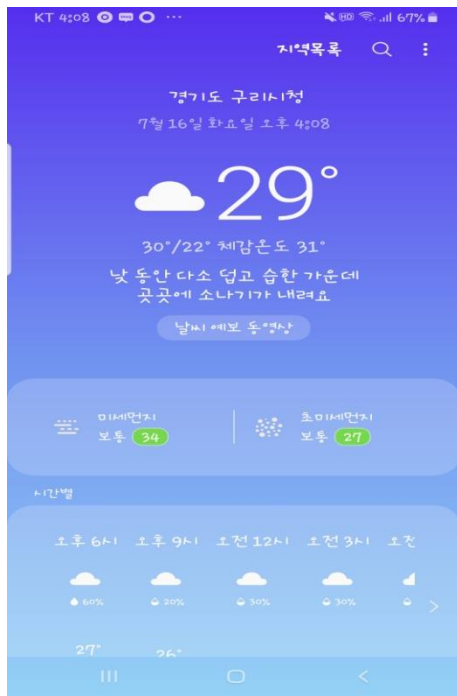
04 분석 결과

05 기대효과

06 한계점

07 Q & A

기존 날씨 Application



공통 정보 : 현재 온도/ 최고 온도/ 최저 온도/ 강수 정보/ 맑음 정도

Why "Wheather me"?

01 주제 선정 배경

Overall

Existing Applications

Necessity

Weather me

02 데이터 정의

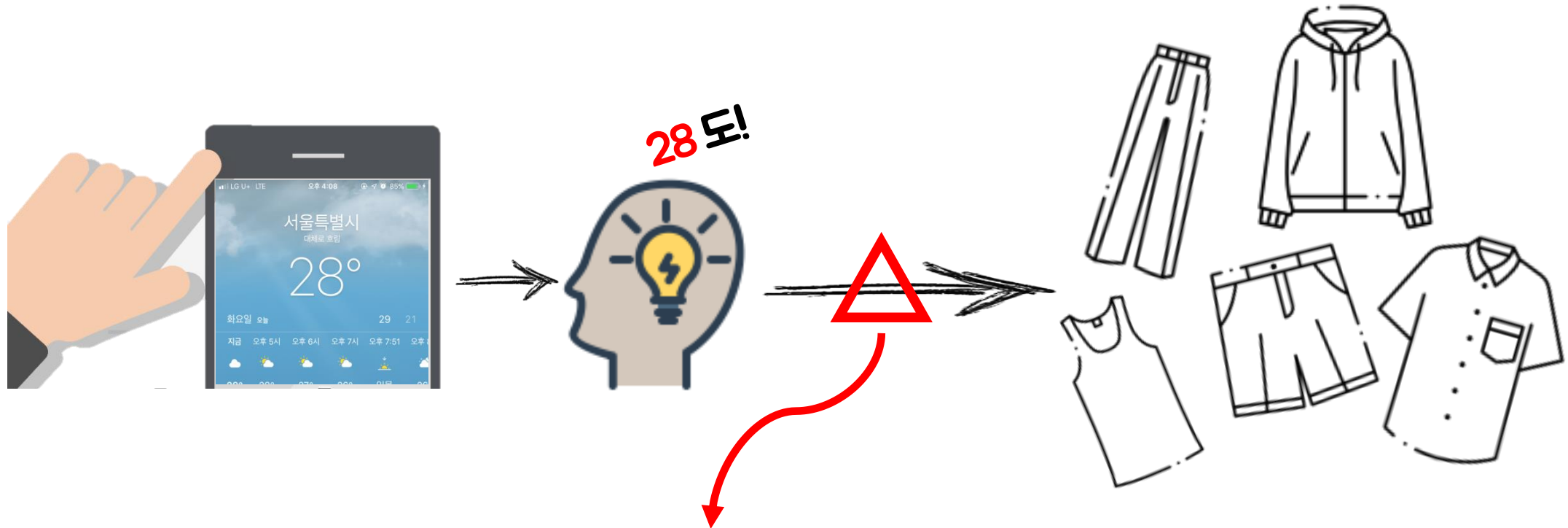
03 데이터 처리 방안

04 분석 결과

05 기대효과

06 한계점

07 Q & A

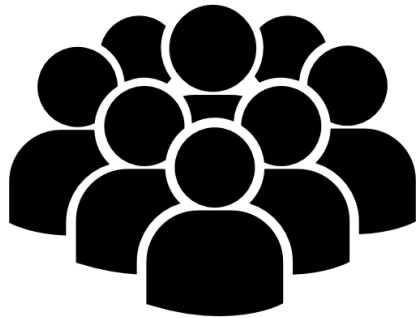


단순한 수치 정보만 가지고 날씨에 맞는 옷을 선택하기엔
큰 일교차, 체감 온도, 더딘 날씨 감각 등의 변수가 존재하기 때문에
적절한 가이드라인 제공이 필요!

Q. Weather Me 사용 의향이 있으신가요?

Weather Me의 기능을 설명하고, 사용 의사를 총 102명의 사람들에게 설문

YES!



73% (74 명)

NO



27% (28 명)

01 주제 선정 배경

Overall

Existing Applications

Necessity

Weather me

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대효과

06 한계점

07 Q & A

What is "Weather me"?

Weather Me

01 주제 선정 배경

Overall

Existing Applications

Necessity

Weather me

02 데이터 정의

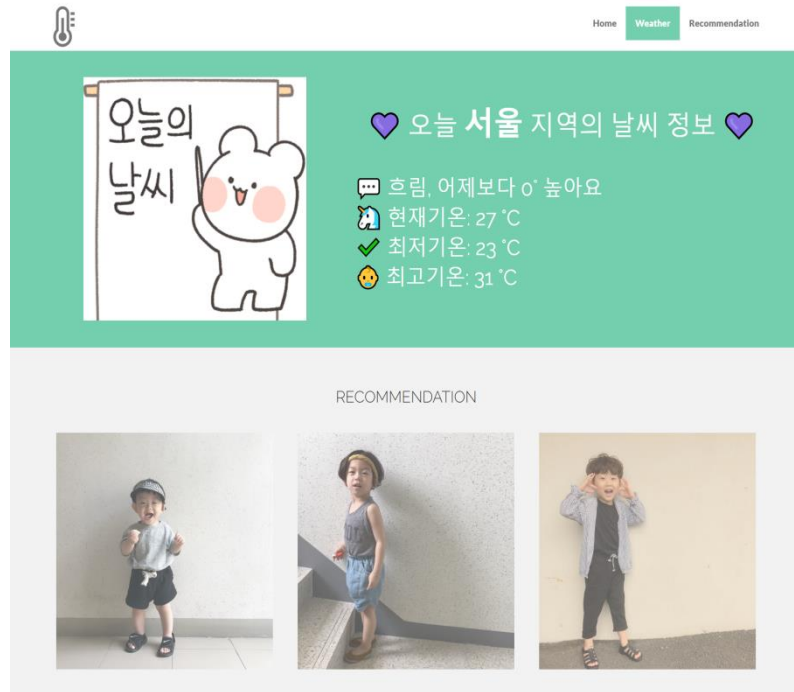
03 데이터 처리 방안

04 분석 결과

05 기대효과

06 한계점

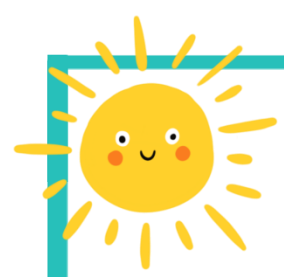
07 Q & A



"Weather Me"

#위치기반 #날씨
아이옷 스타일을 추천해줍니다

아이의 패션이나 건강을 중요시하는 부모님들을 타겟으로
빅데이터를 활용한 분야가 많아지는 요즘, 위치정보에 따라 날씨데이터와 인스타 크롤링으로 얻어온 사진들을 분석함으로써 날씨정보, 아이 옷 선택 및 최신 패션 트렌드 파악에 도움을 준다.



2. 데이터 정의

기상 데이터

01 주제 선정 배경

02 데이터 정의

기상 데이터

이미지 데이터

03 데이터 처리 방안

04 결과 분석

05 기대효과

06 한계점

07 Q & A

기온_20180601_20190531_강원도_원주_...	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_경기도_수원_...	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_경기도_수원_...	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_경기도_수원_...	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_경기도_양평_...	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_경북_구미_구미	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_광주_광주	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_대구_대구	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_대전_대전	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_부산_부산	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_서울_서울	2019-07-08 오후 7...	Microsoft Excel 쉼...
기온_20180601_20190531_울산_울산	2019-07-08 오후 7...	Microsoft Excel 쉼...

일시	월	일	월시기	평균기온_℃	최저기온_℃	최고기온_℃
2018.6.1	6	1	상	23.8	17.5	30.2
2018.6.2	6	2	상	23.4	17.6	30.1
2018.6.3	6	3	상	24	16.9	30.8
2018.6.4	6	4	상	22.6	18.9	27.5
2018.6.5	6	5	상	23.7	17.7	29.2
2018.6.6	6	6	상	23	18.4	27.9
2018.6.7	6	7	상	22.2	17.8	29.4
2018.6.8	6	8	상	21.9	18.5	27
2018.6.9	6	9	상	23.7	18.9	30.1
2018.6.10	6	10	상	21.3	18.3	24.9
2018.6.11	6	11	중	20.2	18	23.9
2018.6.12	6	12	중	21.3	16.6	27.1
2018.6.13	6	13	중	22	18.3	27
2018.6.14	6	14	중	21.6	18.8	25.1
2018.6.15	6	15	중	21.7	17.2	27
2018.6.16	6	16	중	22.5	16.3	28.7

출처: 기상청 기상자료개방포털

지역 & 날짜 별 평균 기온, 최고 기온, 최저 기온,
강수량, 풍량, 미세먼지 수치 etc

지역 정보 (15 도시)

날짜 정보 (2018.05 - 2019.06)

평균 기온

최저 기온

최고 기온



Instagram

인스타그램의 이미지 데이터와 연결

이미지 데이터

01 주제 선정 배경

02 데이터 정의

기상 데이터

이미지 데이터

03 데이터 처리 방안

04 결과 분석

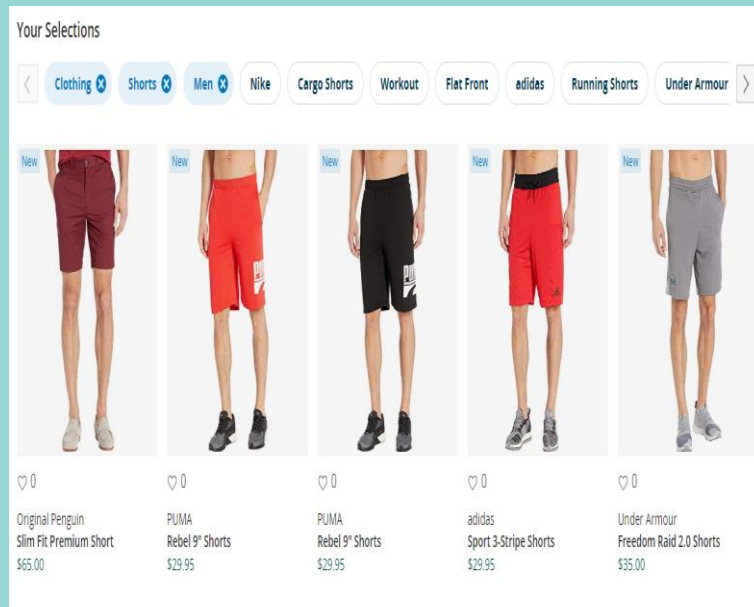
05 기대효과

06 한계점

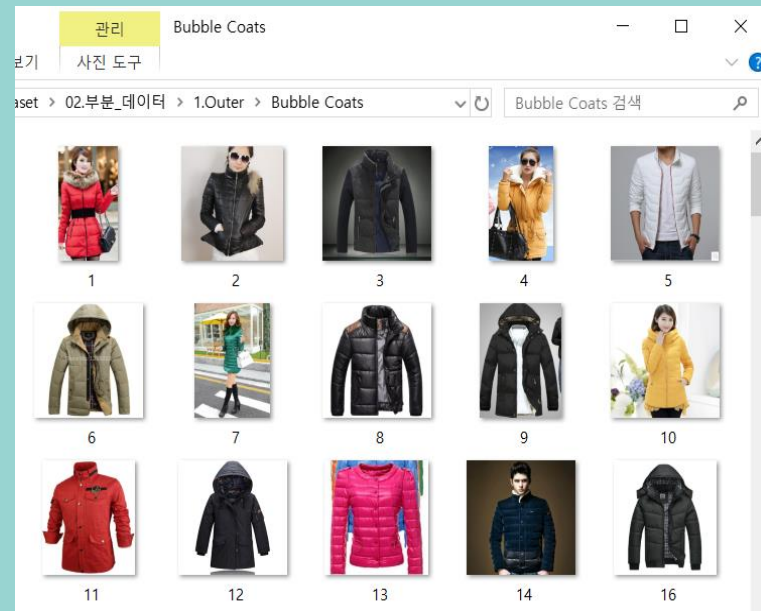
07 Q & A

1. **캐글**에서 제공하는 데이터

2. 부족한 데이터를 **쇼핑몰**에서 크롤링



쇼핑몰 사진



캐글 데이터



이미지 데이터

01 주제 선정 배경

02 데이터 정의

기상 데이터

이미지 데이터

03 데이터 처리 방안

04 결과 분석

05 기대효과

06 한계점

07 Q & A

크롤링을 통한 인스타그램 이미지 데이터 확보

아이디와
지역 정보

사진

남자 유아
+

높은 Daily look 게시물 업로드 수

35개의 인스타 계정
총 6205 개의 사진

날짜 정보

이미지 데이터와 기상 데이터의 연결

01 주제 선정 배경

02 데이터 정의

기상 데이터

이미지 데이터

03 데이터 처리 방안

04 결과 분석

05 기대효과

06 한계점

07 Q & A

Instagram



아이디 : limeunzung

지역 : Golden Gate Park

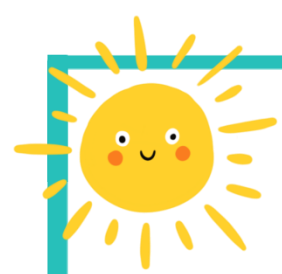
날짜 : 2017 /05/01

최고 온도 : 26

최저 온도 : 20

평균 온도 : 23

지역 정보를 기준으로
이미지 데이터와 기상데이터를 연결시키면
사진 **한 장 당 6개의 정보**를 갖게 된다.



Python, BeautifulSoup, Crawling,
KERAS, CNN



3. 데이터 처리 방안

날씨 분류하기

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

날씨 그룹
데이터 처리

04 분석 결과

05 기대효과

06 한계점

07 Q & A

1. 관측지점 15개의 날씨 데이터를 병합 (5475 열)
2. $\text{diff} = \text{최고기온} - \text{최저기온}$
diff와 평균 diff의 차이를 비교해 한 가지 cluster를 2개로 나눔
3. 총 16등급으로 최종 날씨 범주화



Cluster Num	평균 온도	일교차 차이
1	4도 이하	이하
2	4도 이하	이상
3	4도 초과 8도 이하	이하
4	4도 초과 8도 이하	이상
5	8도 초과 11도 이하	이하
6	8도 초과 11도 이하	이상
7	11도 초과 16도 이하	이하
8	11도 초과 16도 이하	이상
9	16도 초과 19도 이하	이하
10	16도 초과 19도 이하	이상
11	19도 초과 22도 이하	이하
12	19도 초과 22도 이하	이상
13	22도 초과 27도 이하	이하
14	22도 초과 27도 이하	이상
15	27도 초과	이하
16	27도 초과	이상

이미지 전처리 (1)

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

날씨 그룹

데이터 처리

04 분석 결과

05 기대효과

06 한계점

07 Q & A

1. 앞에서 지역 정보를 통해 기상 데이터와 이미지 데이터를 연결
2. 인스타그램에서 크롤링한 **이미지 데이터도** 기상 데이터 그룹과 같이 **16개의 폴더로 나누어 날씨별로 저장**

Cluster Num	평균 온도	일교차 차이
1	4도 이하	이하
2	4도 이하	이상
3	4도 초과 8도 이하	이하
4	4도 초과 8도 이하	이상
5	8도 초과 11도 이하	이하
6	8도 초과 11도 이하	이상
7	11도 초과 16도 이하	이하
8	11도 초과 16도 이하	이상
9	16도 초과 19도 이하	이하
10	16도 초과 19도 이하	이상
11	19도 초과 22도 이하	이하
12	19도 초과 22도 이하	이상
13	22도 초과 27도 이하	이하
14	22도 초과 27도 이하	이상
15	27도 초과	이하
16	27도 초과	이상



보기		
dataset > 01.인풋_데이터 > 안자른것 > insta		
insta 검색		
이름	수정한 날짜	유형
1	2019-07-13 오후 1...	파일 폴더
2	2019-07-13 오후 1...	파일 폴더
3	2019-07-13 오후 1...	파일 폴더
4	2019-07-13 오후 1...	파일 폴더
5	2019-07-13 오후 1...	파일 폴더
6	2019-07-13 오후 1...	파일 폴더
7	2019-07-13 오후 1...	파일 폴더
8	2019-07-13 오후 1...	파일 폴더
9	2019-07-13 오후 1...	파일 폴더
10	2019-07-13 오후 1...	파일 폴더
11	2019-07-13 오후 1...	파일 폴더
12	2019-07-13 오후 1...	파일 폴더
13	2019-07-13 오후 1...	파일 폴더
14	2019-07-13 오후 1...	파일 폴더
15	2019-07-13 오후 1...	파일 폴더
16	2019-07-13 오후 1...	파일 폴더

이미지 전처리 (2)

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

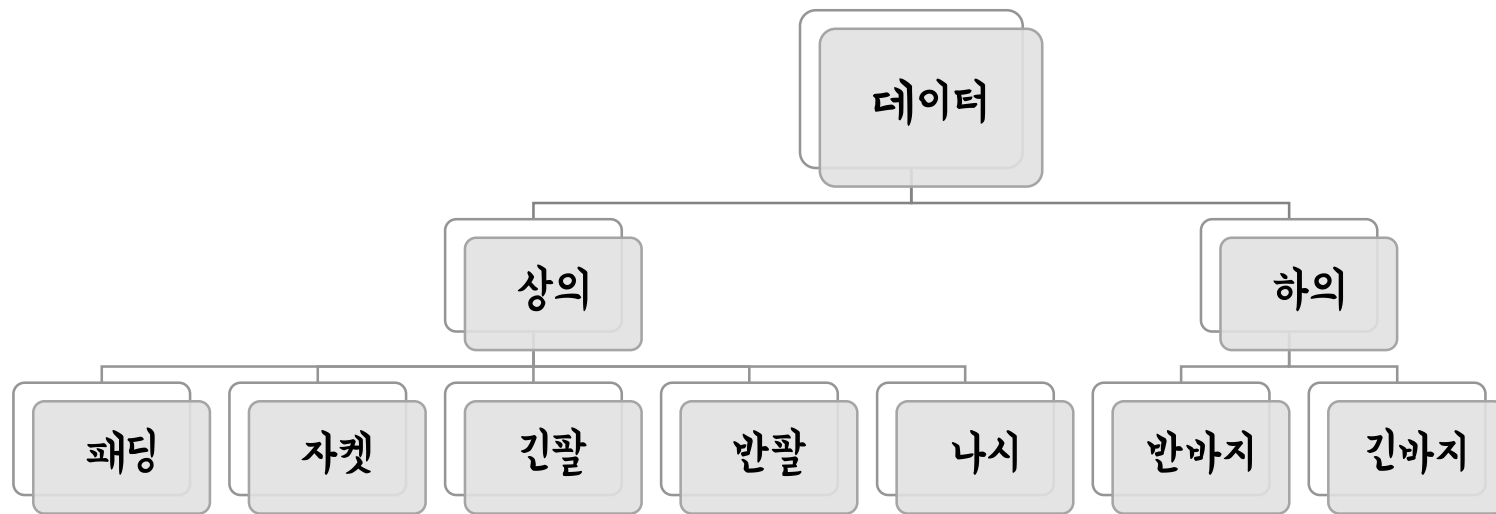
날씨 그룹
데이터 처리

04 분석 결과

05 기대효과

06 한계점

07 Q & A



전신을 볼 수 있는 사진 외 사진 삭제



사진 속 아이의 옷 종류를 기준으로 사진을 **카테고리화**



상의는 5개 **라벨**로, **하의**는 2개 **라벨**로 분류

이미지 전처리 (3)

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

날씨 그룹

데이터 처리

04 분석 결과

05 기대효과

06 한계점

07 Q & A

비정형화(자료부족, 모호성) 되어있는 인스타그램 사진 데이터 삭제

기존 데이터

조기, 패딩조기, 레이어드 룩, 7부바지
등 애매한 카테고리의 데이터 포함



인스타그램 35개 계정
크롤링 이미지 6205개

CLEAN 1 DATA

1차 데이터 제거

상의	하의
나시: 223개	
반팔: 628개	긴바지: 675개
긴팔: 777개	반바지: 266개
자켓: 457개	
패딩: 286개	

삭제 목록

: 패딩조기, 베스트, 오버롤, 기저귀, 앞
아서 찍은 사진

CLEAN 2 DATA

1차 데이터 제거 + 2차 데이터 제거

상의	하의
나시: 145개	
반팔: 429개	긴바지: 421개
긴팔: 541개	반바지: 138개
자켓: 398개	
패딩: 275개	

삭제 목록

: 뒷모습, 짧은 바지+ 레깅스, 한쪽다리,
비딱한 자세, 옆모습

Rotation 활용

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

날씨 그룹
데이터 처리

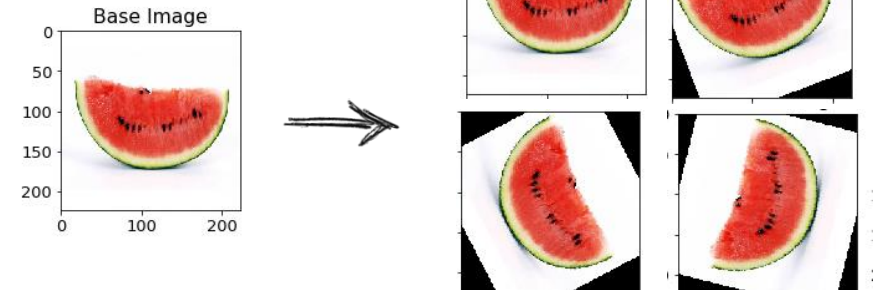
04 분석 결과

05 기대효과

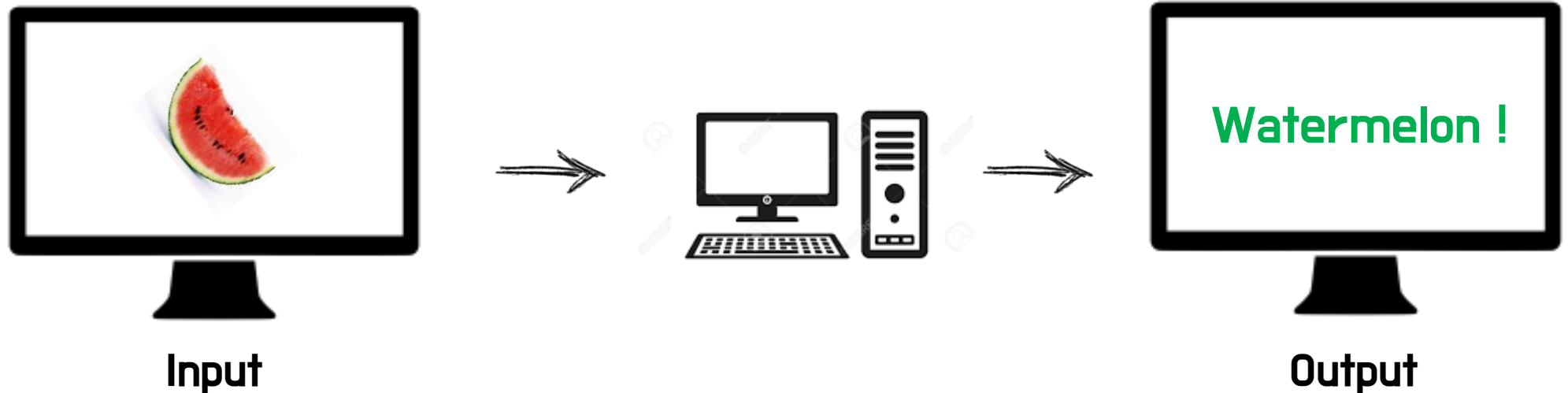
06 한계점

07 Q & A

```
data = ImageDataGenerator(rescale=1./255,  
                           rotation_range=25,  
                           width_shift_range=0.1,  
                           height_shift_range=0.1,  
                           shear_range=0.3,  
                           zoom_range=0.2)
```



이미지 데이터의 수를 증가시키고 다양한 각도에서의 이미지 인식 가능하게 하여 **오버 피팅 방지**



Rotation 활용

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

날씨 그룹

데이터 처리

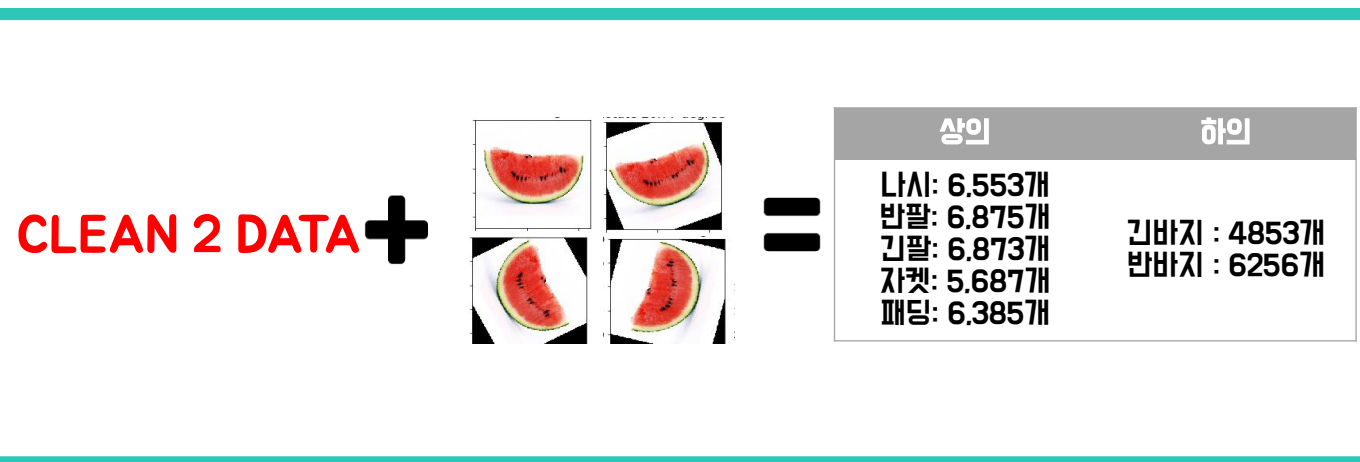
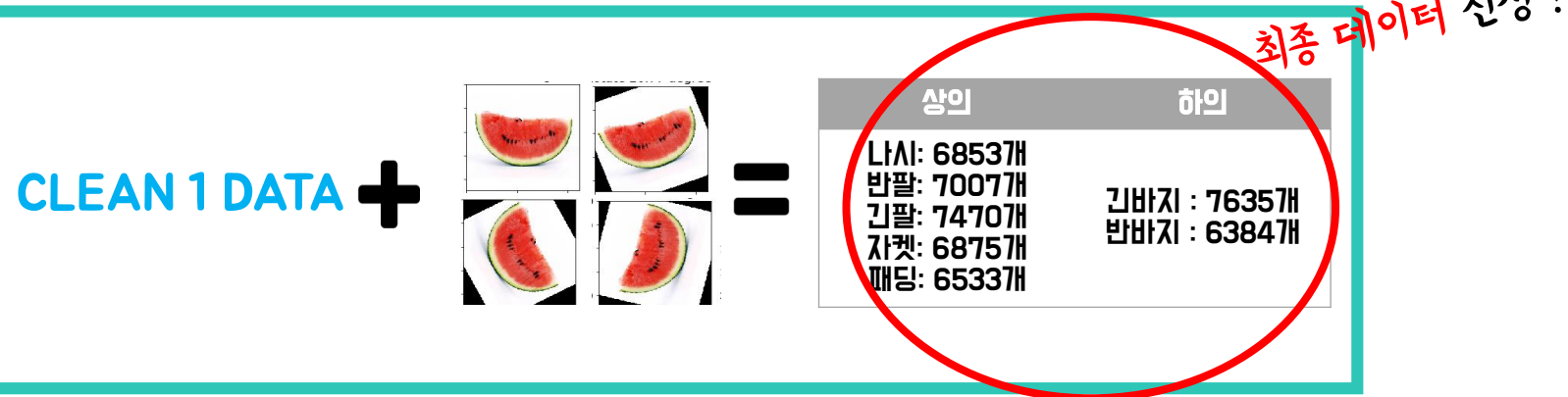
04 분석 결과

05 기대효과

06 한계점

07 Q & A

ROTATION활용 전에는 CLEAN2 DATA의 결과가 좋았지만, **ROTATION**후 정확도는 **CLEAN 1 DATA**가 더 높게 측정됨

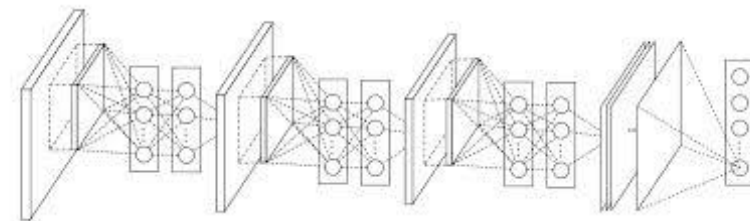


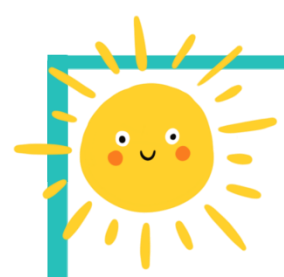
〈 기본 설정 〉

- 정방향 컬러 이미지 사용
- 입력 이미지 의 크기는 32X32, 64X64, 128X128으로 변경하여 실험

〈 레이어의 구성 〉

- 세 개의 컨볼루션 레이어
- 활성화 함수 (activation function)는 ReLU를 사용
- 두 개의 max pooling 레이어
- softmax 분류를 사용한 결과 레이어
- 세 번의 drop out (0.25, 0.25, 0.5)
- batch size:32 , epoch: 50,100,200





4. 분석 결과

쇼핑몰 데이터
① Zappos.
② Kaggle dataset

상의			하의		
image size	epoch	accuracy	image size	epoch	accuracy
32	50	0.8643	32	50	0.9932
	100	0.8392		100	0.9932
	200	-		200	0.9966
64	50	0.8641	64	50	0.6488
	100	-		100	0.9864
	200	-		200	0.623
128	50	0.6341	128	50	0.621
	100	-		100	0.643
	200	-		200	0.432

모델에 **인스타 데이터** 적용시.
예측력 부족

➡ **인스타 데이터** 사용하여 모델링 실시

모델 구성

- Epoch 50
- 합성곱층 3
- 최대 풀링 2
- 활성화함수 relu, softmax
- Dropout 3 (0.25, 0.25, 0.5)



데이터 선정

- 상의
clean1-rotate 이미지32
- 하의
clean2-rotate 이미지32

- 01 주제 선정 배경
- 02 데이터 정의
- 03 데이터 처리 방안
- 04 결과 분석
 - 실패 사례 소개 및 분석
 - 성공 사례 소개 및 분석
- 05 기대효과
- 06 한계점
- 07 Q & A

상의 clean2-rotate 이미지32 -> **clean1-rotate 이미지32** 변경된 이유??

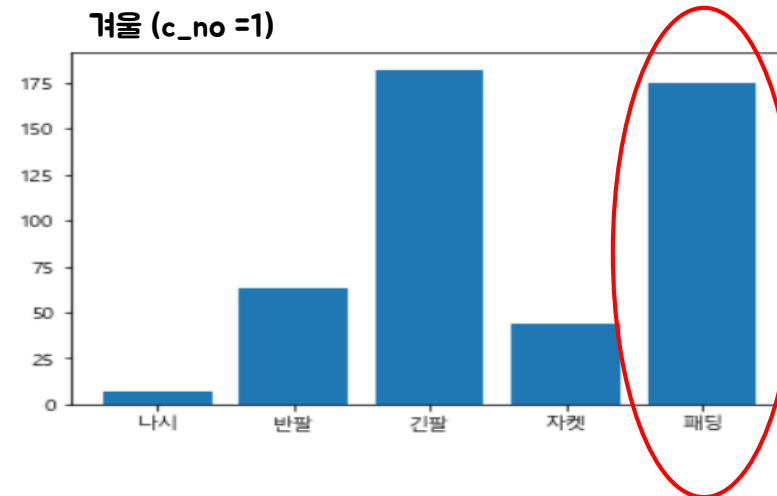
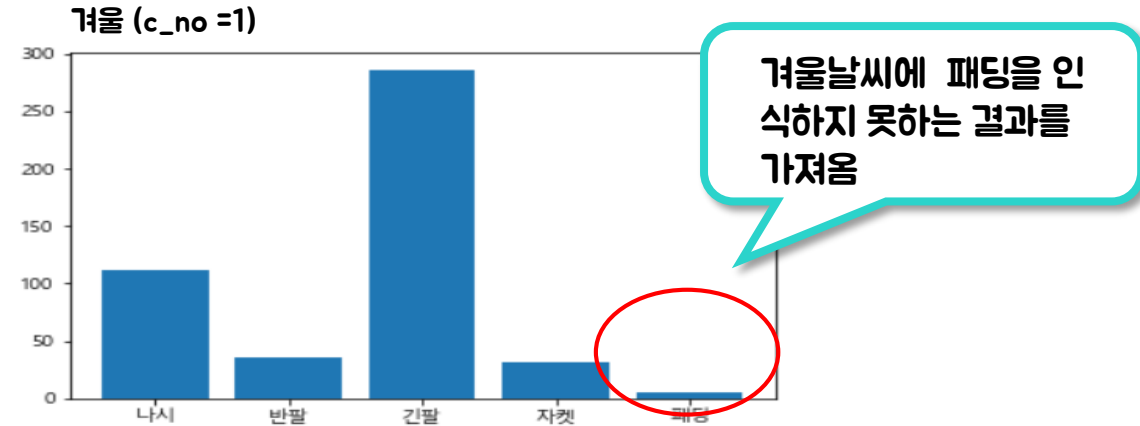
이미지 32

Clean2- rotate	
상의	하의
loss = 0.398 accuracy = 0.843	loss = 0.365 accuracy = 0.956



이미지 32

Clean1- rotate	
상의	하의
loss = 1.54 accuracy = 0.8	loss = 0.221 accuracy = 0.971



01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

04 결과 분석

실패 사례 소개 및 분석

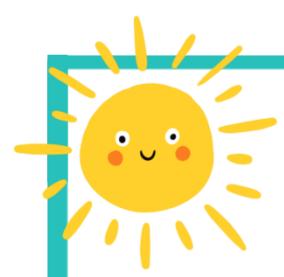
성공 사례 소개 및 분석

05 기대효과

06 한계점

07 Q & A

날씨 그룹			상의				하의			
c_no	평균온도	일교차차이	나시	반팔	긴팔	자켓	패딩	반바지	긴바지	
1	4도 이하	이하	긴팔, 긴바지	1%	39%	9%	37%	21%	79%	
2	4도 이하	이상	패딩, 긴바지	1%	29%	19%	36%	17%	83%	
3	4도 초과 8도 이하	이하	자켓, 긴바지	1%	4%	34%	30%	17%	83%	
4	4도 초과 9도 이하	이상	긴팔, 긴바지	3%	27%	15%	26%	31%	69%	
5	8도 초과 11도 이하	이하	반팔, 긴바지	4%	32%	9%	25%	35%	65%	
6	8도 초과 11도 이하	이상	패딩, 긴바지	5%	23%	11%	37%	20%	80%	
7	11도 초과 16도 이하	이하	반팔, 긴바지	0%	54%	8%	14%	25%	75%	
8	11도 초과 16도 이하	이상	반팔, 긴바지	4%	42%	12%	21%	37%	63%	
9	16도 초과 19도 이하	이하	긴팔, 긴바지	2%	13%	32%	8%	32%	68%	
10	16도 초과 19도 이하	이상	긴팔, 긴바지	3%	29%	11%	26%	48%	52%	
11	19도 초과 22도 이하	이하	반팔, 긴바지	4%	39%	10%	12%	41%	59%	
12	19도 초과 22도 이하	이상	반팔, 반바지	5%	48%	16%	6%	74%	26%	
13	22도 초과 27도 이하	이하	반팔, 반바지	12%	59%	8%	6%	69%	31%	
14	22도 초과 27도 이하	이상	반팔, 반바지	10%	64%	3%	5%	87%	13%	
15	27도 초과	이하	반팔, 반바지	21%	56%	3%	4%	86%	14%	
16	27도 초과	이상	반팔, 반바지	26%	59%	0%	2%	86%	14%	



5. 기대효과



Weather me

날씨에 민감한 우리 아이에게 알맞은 옷을 입혀주세요.
날씨기반 아이옷 추천서비스, 웨더미

click





01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대 효과

Website

App

Style

06 한계점

07 Q&A



♥ 오늘 서울 지역의 날씨 정보 ♥

💬 흐림, 어제보다 0° 높아요

🦄 현재기온: 27 °C

✓ 최저기온: 23 °C

😓 최고기온: 31 °C

RECOMMENDATION



Application

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대 효과

Website

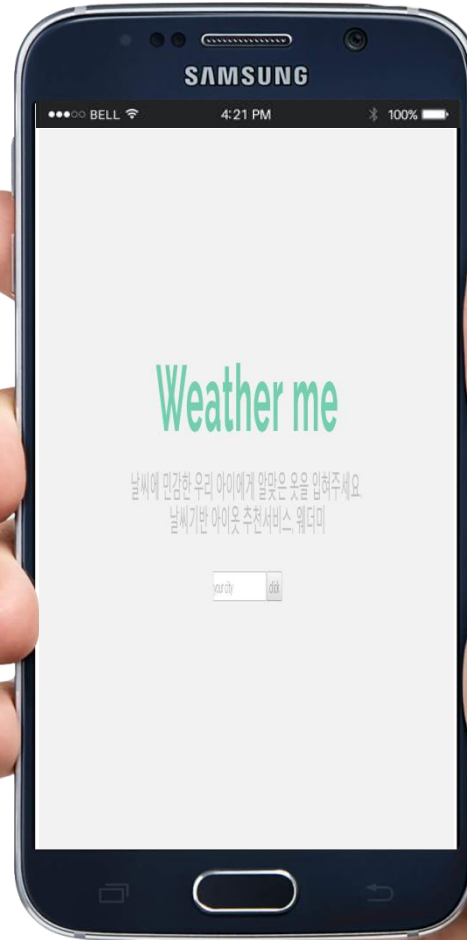
App

Style

06 한계점

07 Q&A

현재 날씨, 일교차, 강수, 미세먼지
등과 같은 다양한 날씨확인과 동시
에 알맞은 아이 옷 추천 서비스 이용
가능



지그재그, 브랜디, 에이블리 등의 어플리케
이션 기능과 같이 여러 쇼핑몰들을 한곳에
모아 쇼핑의 편의성을 제공

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대 효과

Website

App

Style

06 한계점

07 Q&A

인스타그램 API 연동 회원가입을 기반으로 하는 스타일 추천 서비스 운영

기술 연동과 서비스 고도화로 Weather Me를 이용하는 이용자의 만족도를 높이는 전략

여러 아동 쇼핑몰 제휴 + 빅데이터 분석

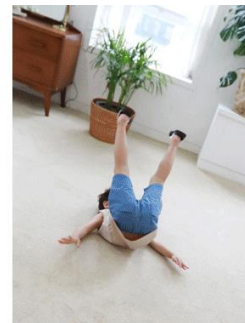
① 인스타그램 API 연동 회원가입제도 도입

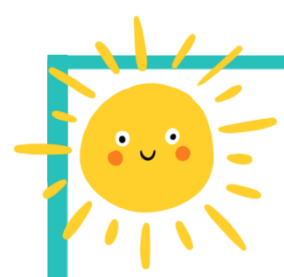
② 계정에 업로드된 이미지 데이터를 수집, 분석 후 스타일 책정

③ 책정된 스타일에 알맞은 스타일 추천



Recommendation





6. 한계점

한계 및 보완점

01 주제 선정 배경

02 데이터 정의

03 데이터 처리 방안

04 분석 결과

05 기대효과

06 한계점

07 Q & A

1. 16개 날씨 그룹별 옷의 특징이 뚜렷하지 않음

- 같은 기온의 날씨지만, 일교차가 크면 아이들 옷 스타일이 다를 것으로 가정했으나, 사실상은 2개 이상의 그룹이 대부분 반팔을 입는 것으로 나타나 각 그룹당 특성을 뽑아내기 어려웠음
- 날씨 그룹의 수(16)와 인스타 이미지의 카테고리 수(상의5,하의2) 간의 괴리

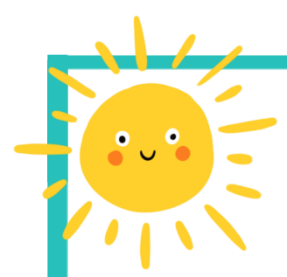
2. 유효한 인스타 이미지 수 대량 확보 필요

- CNN 상의, 하의 카테고리 다양화, 세분화

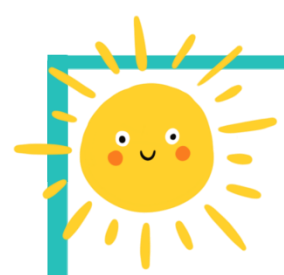
3. 상 하의 자동 추출 미비 (YOLO, OpenCV, google Vision API)

4. CNN 모델 다양화, 고도화 지속적인 시도 필요

5. 날씨, 날씨가 옷들과의 직접적인 연관관계 또는 유형화 분석 시도 필요



Q & A



감사합니다

