



어그로(주목도) 높은 유튜브
썸네일 이미지 분류 AI

어썸 판별기

3조

박종호 유웅기 이주현 이혜진

목차

프로젝트 개요

1. 시장 현황
2. 프로젝트 기획 의도
3. 해결 방안

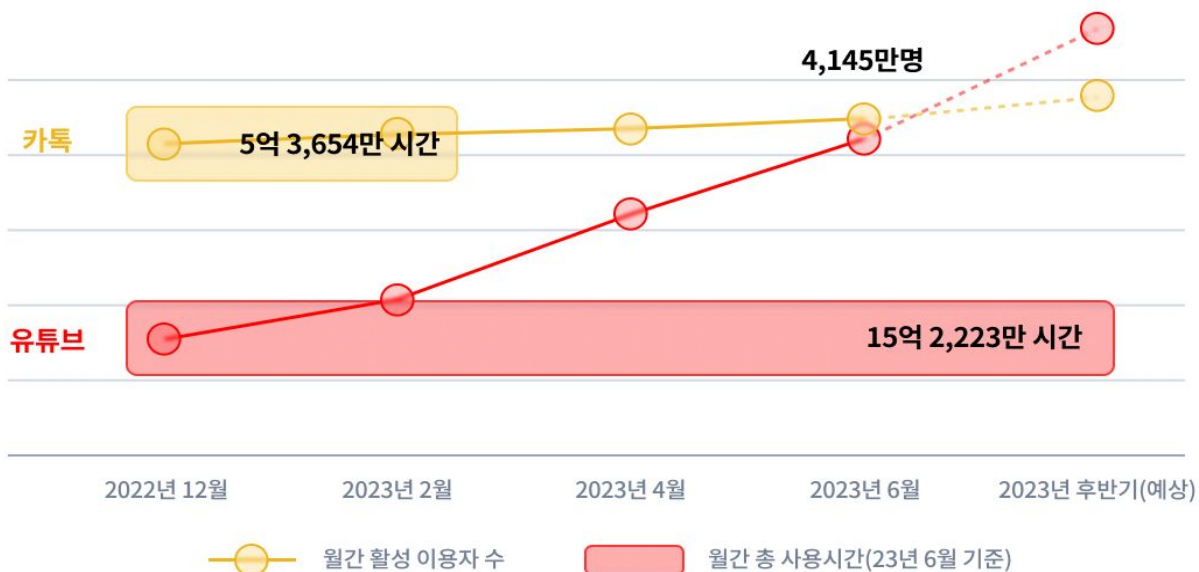
프로젝트 수행

1. 프로젝트 팀 구성 및 역할
2. 프로젝트 수행 절차 및 방법
3. 프로젝트 수행 결과
4. 자체 평가 의견

시장 현황 - 국민앱 Youtube

.....

카카오톡과 유튜브의 월간 이용자 수, 총 사용시간 비교



월간 이용자수와 사용시간 카카오톡을 앞지름

*출처: applovin

시장 현황 - MZ는 소비자이자 생산자

.....



(Base : 전체, n=800, 단위 : %)



유튜브 영상 콘텐츠 생산 주요 목적 TOP3

(Base : 유튜브 생산 경험자, n=211, 단위 : %)

①

친구·지인들과
공유하기 위해

23.7%

②

일상을 기록·
추억하기 위해

23.2%

③

숙제·과제·업무 등
미션을 위해

18.5%

“

유튜버는
전문적인
직업이 될 수 있다

68.3%

“

기회가 된다면
영상 제작에 필요한
전문 기술을 배우고 싶다

65.5%

“

기회가 된다면
유튜버를 준비하거나
유튜버를 한 번쯤 해볼 것이다

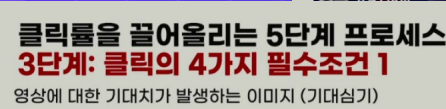
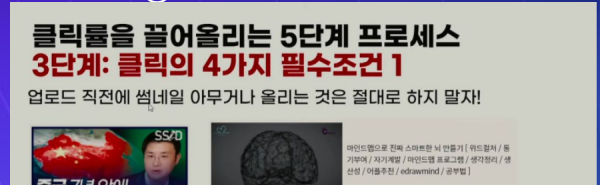
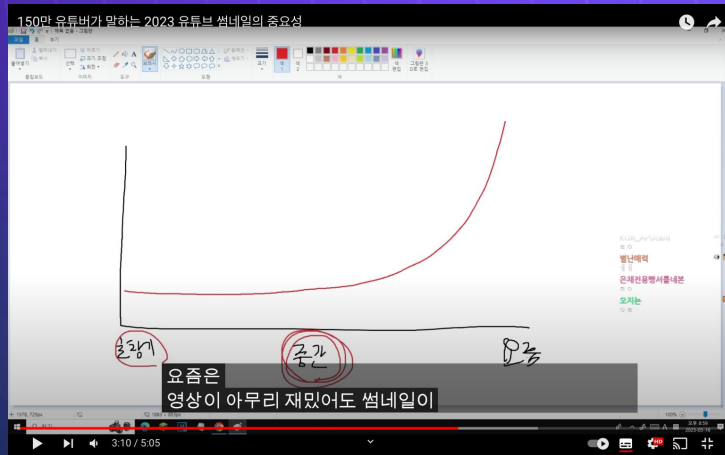
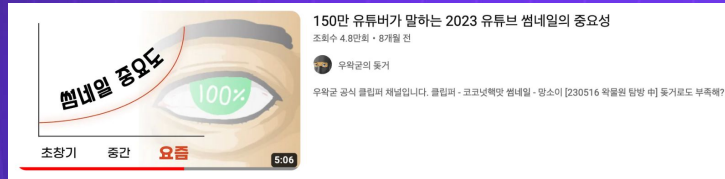
57.6%

유튜버는 전문 직업, 기회가 된다면 한 번쯤 해보고 싶어

*출처: 대학내일 20대 연구소

기획 의도 - 썸네일의 중요성

.....



레드오션에서 살아남으려면?

기획 의도 - 유튜브 썸네일 정책

.....

클릭률 데이터와 관련하여 해서는 안 되는 작업은 무엇인가요?



- 충분한 데이터 없이 결정하는 일. 상당한 노출수가 발생한 후에 클릭률 데이터를 살펴보시기 바랍니다. 업로드하고 나서 바로 클릭률을 확인하지 마세요.
- 클릭률에서 발생한 약간의 변화를 개선하는 일. 클릭률에 작은 변동이 생기는 일은 일반적이며 즉각적인 조치가 필요하지 않습니다. 클릭률의 변화가 통계적으로 유의미하다면 개선이 필요할 수 있습니다.
- 같은 동영상에 여러 썸네일이나 제목을 테스트하는 일. 각 동영상을 동일한 시청자층이 보도록 하기는 어렵습니다. 클릭률의 차이는 제목이나 썸네일보다는 트래픽 소스 때문일 수 있습니다.

사용자 대상으로 테스트하지 말라는 유튜브 정책
하지만 유튜버들은 영상 업로드 후 데이터를 확인해야 하는 실상

기획 의도 - 썸네일의 중요성

.....

Psychology & Cognition
36 (7), 1351-1359
© 2008 The Authors
Journal compilation © 2008 British Psychological Society

The picture superiority effect in associative recognition

WILLIAM E. HOCKLEY

Wilfrid Laurier University, Waterloo, Ontario, Canada

The picture superiority effect has been well documented in tests of item recognition and recall. The present study shows that the picture superiority effect extends to associative recognition. In three experiments, students studied lists consisting of random pairs of concrete words and pairs of line drawings; then they discriminated between intact (old) and rearranged (new) pairs of words and pictures at test. The discrimination advantage for pictures over words was seen in a greater hit rate for intact picture pairs, but there was no difference in the false alarm rates for the two types of stimuli. That is, there was no mirror effect. The same pattern of results was found when the test pairs consisted of the verbal labels of the pictures shown at study (Experiment 4), indicating that the hit rate advantage for picture pairs represents an encoding benefit. The results have implications for theories of the picture superiority effect and models of associative recognition.

3

Table 3
Mean Hit Rates (HR), False Alarm Rates (FAR), and Estimates of d' and c , and the Standard Errors of the Means (SE) for the Word and Picture Pairs Tested in Experiment 3

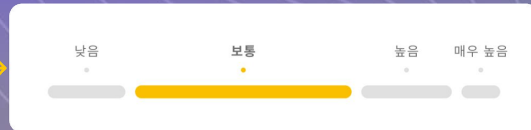
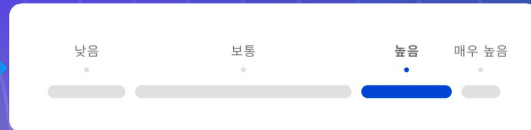
Pair	HR		FAR		d'		c	
	M	SE	M	SE	M	SE	M	SE
Words	.56	.03	.20	.03	1.24	0.17	.43	.07
Pictures	.82	.02	.29	.03	1.71	0.17	-.16	.07

이미지 우월성 효과 :
사람들이 시각적 이미지를 텍스트나 구어보다
더 빠르게 인지하고 더 잘 기억하는 현상

유튜브 어그로 썸네일 판별기

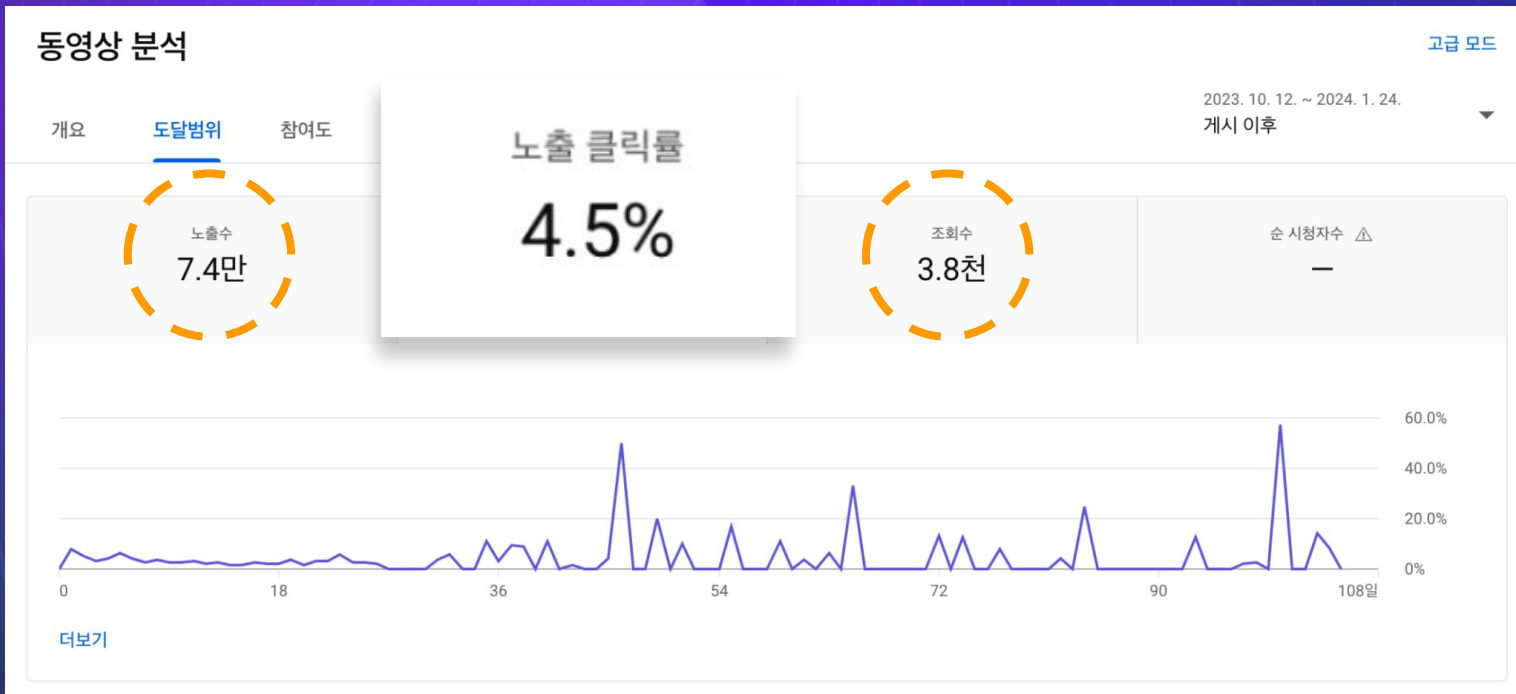
.....

유튜브 크리에이터



분류 기준 기획

.....



노출수는 수집 불가능
대신 구독자수 대비 조회수로 클릭률 높은 영상 예상



Our Team

박종호

- 데이터 전처리
- 데이터 구축
- 모델 학습

유웅기

- 서비스 기획 및 설계
- 모델 학습

이주현

- 팀장
- 데이터 수집
- 모델 학습

이혜진

- 모델 구축
- 모델 학습

데이터 수집 - 구독자수 Top 채널

.....

TOP 300	게임	BJ/인물/연예인	음악/댄스/가수	TV/방송	음식/요리/레시피	패션/미용	뉴스/정치/사회	취미/라이프	IT/기술/컴퓨터
회사/오피셜	교육/강의	영화/만화/애니	키즈/어린이	애완/반려동물	스포츠/운동	국내/해외/여행	자동차	주식/경제/부동산	미분류
전체순위	구독자순	동영상순	재생순	조회순	이름	카테고리순	Random		

172 장박문	277 MBN MBN Entertainment	214 GIRLS GENERATION	155 FC BAYERN MUNICH	255 Kep1er	46 SERIE A	278 창작맨	161 EVANS	47 Mnet TV	148 워크맨 Workman
92 PinkPopTOY	55 ONE Championship	32 DubuPopTOY	230 KQ ENTERTAINMENT	167 ATEEZ	키즈/어린이 카테고리	54 THE BEATLES	291 RISABAE	106 Louis Vuitton	260 PLAYLIST ORIGINALS 플레이리스트 오리진얼
201 Adventure Story	189 3D바스타	259 freezy	89 PONY Syndrome	101 Red Velvet	41 M2	281 [onct] MUSIC 채널 NCT MUSIC	57 JTBC enter	296 [wonderkids TV] 원더키즈 TV	102 KBS Entertain
BJ/인물/연예인 BJ/인물/연예인 카테고리	91 Maytree	137 유희열의 스케치북	145 tvN	236 1-MINUTE 1분 요리/초대미팅	224 나뵤Nado	240 OfficialGDRAGON	149 KBS DRAMA	289 올리비엔	53 KBS KPOP KBS Kpop
178 이영준 위클리	242 ONE PIECE	187 리얼리티WORLD	235 EBS Culture EBS Culture (EBS 2FM)	218 조선대학교	129 [onct] DAILY 채널 NCT DAILY	207 The Wiggles Kids Songs and Nursery Rhymes	86 소라카렌	244 Zip - Nursery Rhymes	150 TV PP TVPeople

유튜버 랭킹 사이트 내 카테고리별 랭킹 참고해 총 1000개 채널 수집

데이터 수집 - 키워드별 유튜브 검색 결과

.....

YouTube Premium KR

먹방

연애보다 밥 먼저 - 매콤 명란 파스타 & 거울 호떡 편
조회수 93,931 · 22시간 전
총맛명
오늘도 시청해주셔서 감사드립니다!! [총맛명 스튜디오] 총감독 : 총맛명 / 부감독 : 지카오토 / 제작진행 : 스노유진 / 스토리보드 : 호행 ...
새 동영상 자막

24시간 대패삼겹살 집에서 쉬지않고 고기 구워먹으면 얼마나 먹을까요? 최장시간먹방..
조회수 26,000 · 22시간 전
히밥heebab
항학동맛집 #인생대패 #대패삼겹살맛집 이 영상은 인생대패에서 제작료를 지원 받아 제작된 유료광고 영상입니다. 인생대패 인스타 ...
새 동영상

식탁이랑 화면 짝 채우기루 | 낫솔하기루 EP.13
조회수 21,000 · 22시간 전
뽕뽕하기루
업로드 매주 수요일 저녁 6시 비즈니스 문의 jdbent001@gmail.com 구독 좋아요 댓글 알림설정 #신기루 #김민정 ...
새 동영상

사장님도 못 드신다는 매운 짜꾸미 삼겹살 먹방
조회수 16,000 · 2일 전
맵하니Maphany

홈
Shorts
구독
YouTube Music

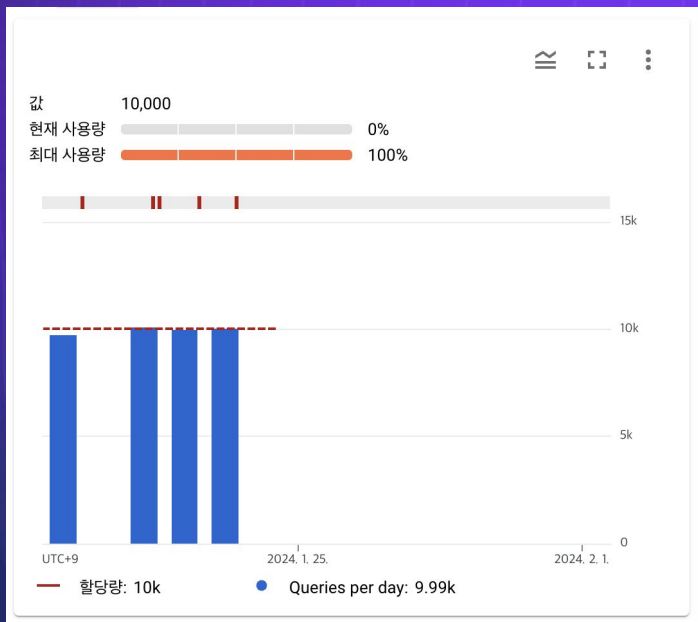
나 >
내 채널
시청 기록
내 동영상
나중에 볼 동영상
오프라인 저장 동영상
더보기

구독
참학맨
슈카월드
Turbo832 TV
육식맨 YOOXICM...
Cami TV Cami

유명 유튜버뿐만 아니라 일반 유튜버 썸네일도 수집하기 위해 키워드 검색
결과 활용

데이터 수집 - YouTube Data API v3

• • • • •



이주현 오후 4:42

CSV 1월 18일 목요일

youtube_video_data_IT.csv

CSV

youtube_video_data_기술.csv

Category	Video Title	View Count	Subscriber Count	Upload Date	Thumbnail URL	Video Duration
기술	이게 된다고? MIT가 선정한 미...	61535	12700	2023-01-24T08:00:02Z	https://i.ytimg.com/vi/...	0:03:00
기술	"옛말로 노가다? 이젠 기술"...MZ...	556257	3990000	2022-12-09T12:23:13Z	https://i.ytimg.com/vi/...	0:03:01
기술	노가다 기술 절대 배우지 말라는...	629335	942000	2022-11-13T04:03:39Z	https://i.ytimg.com/vi/...	0:08:05
기술	1년 3개월만에 타일공 관통이유 ...	37599	2400	2023-06-01T14:37:24Z	https://i.ytimg.com/vi/...	0:06:54
기술	[자막뉴스] 단 4일이면 건물 똑...	530218	4290000	2023-06-05T05:29:46Z	https://i.ytimg.com/vi/...	0:02:16
기술	[자막뉴스] 한국, 전세계 시장 ...	1784559	4290000	2023-09-15T23:00:02Z	https://i.ytimg.com/vi/...	0:01:53
기술	억대연봉 기술직, 뭘해야될지 모...	194746	61900	2021-09-17T08:15:01Z	https://i.ytimg.com/vi/...	0:07:22
기술	세상을 바꿀 미래 혁신 기술 4가...	20445	306000	2020-12-28T08:00:10Z	https://i.ytimg.com/vi/...	0:04:10
기술	고수들이 말하는 삽입의 기술 7...	1251468	351000	2022-08-03T04:52:28Z	https://i.ytimg.com/vi/...	0:09:18
기술	외계인이 만들었다고 밖에 설명할...	867920	1710000	2023-12-10T02:00:07Z	https://i.ytimg.com/vi/...	0:42:46
기술	[자막뉴스] 영하 162도 핵심 기...	393969	2710000	2023-12-31T05:05:59Z	https://i.ytimg.com/vi/...	0:01:53
기술	40대 중반에 퇴직 후 기술배워서...	112223	61900	2022-12-20T08:03:29Z	https://i.ytimg.com/vi/...	0:10:47
기술	해외에서 일용직 잡부로 일하며 ...	163910	3760000	2023-09-27T05:00:16Z	https://i.ytimg.com/vi/...	0:09:17
기술	그녀가 알려주는 최고의 기술	115345	54500	2023-04-18T00:53:33Z	https://i.ytimg.com/vi/...	0:08:01
기술	기술 배워 독립하고 하루 두탕 ...	2418517	272000	2022-04-01T10:06:44Z	https://i.ytimg.com/vi/...	0:14:37
기술	23년 영하 162도 핵심 기술배...	2425200	68700	2023-09-13T00:33:37Z	https://i.ytimg.com/vi/...	0:06:14

Youtube API의 할당량 제한
-> API KEY 약 60개 생성

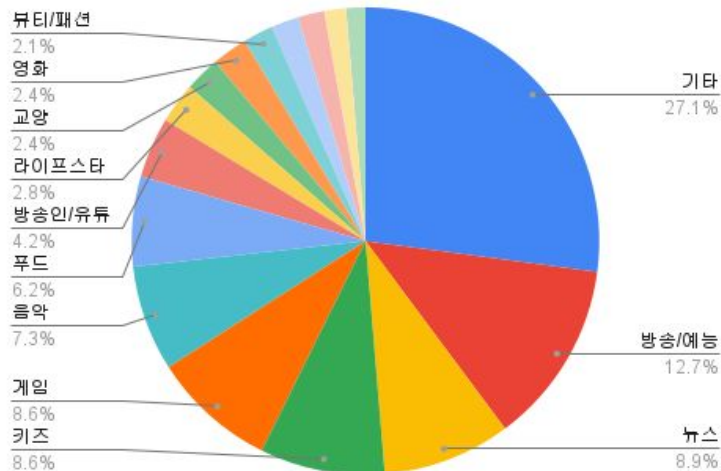
약 32만개 영상의 제목, 조회수, 구독자수, 업로드
날짜, 썸네일 이미지 주소, 영상 길이 수집

데이터 전처리 - 영상 카테고리 분류

.....

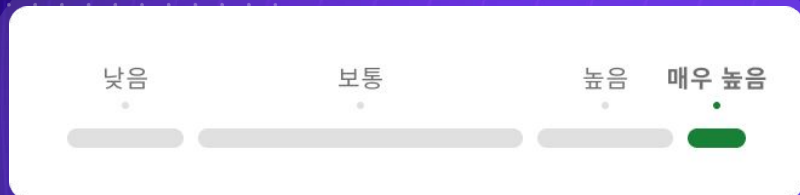
	카테고리	검색 키워드	데이터
1	게임	게임	24,483
2	경제	주식, 경제, 부동산, 회사, 오피셜, 투자, 회사, 경제	4,309
3	교양	교육, 강의, 과학	6,933
4	기타	미분류, 해외	77,016
5	뉴스	뉴스, 정치, 사회	25,182
6	동물	애완, 반려동물, 동물, 강아지, 고양이	5,417
7	라이프스타일	취미, 라이프, 국내, 해외, 여행, 브이로그, 노하우, 캠핑, 취미	7,923
8	방송/예능	TV, 방송, 예능, 코메디, 개그, 유머	36,120
9	방송인/유튜버	BJ, 인물, 연예인, 버튜버, 인물	11,894
10	뷰티/패션	패션, 미용, 스니커즈, 룩북, 메이크업	5,976
11	스포츠	스포츠, 운동, 홈트레이닝, 헬스	3,787
12	영화	영화, 만화, 애니, 애니메이션	6,933
13	음악	음악, 댄스, 가수, 커버 댄스, 노래 커버	20,786
14	키즈	키즈, 어린이	24,521
15	테크	IT, 기술, 컴퓨터, 자동차	5,146
16	푸드	음식, 요리, 레시피, 먹방, 베이킹, 요리	17,687

카테고리별 데이터수



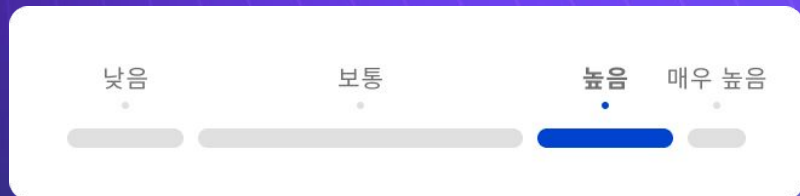
데이터 전처리 - 썸네일 라벨 분류

.....



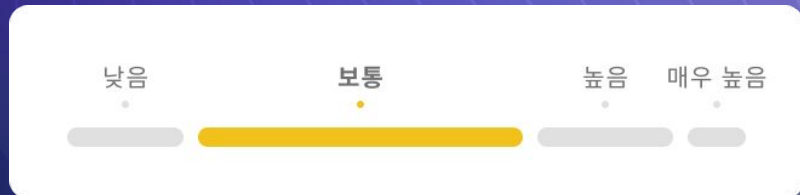
썸네일 주목도 '매우 높음'

- 2021년~2023년 영상 중 조회수 **10,000** 이상
- 구독자수 / 조회수 **0.3** 이상 ('기타' 기준, 카테고리별 다름)
- 총 **62,084** 개 썸네일 이미지 (**21.7%**)



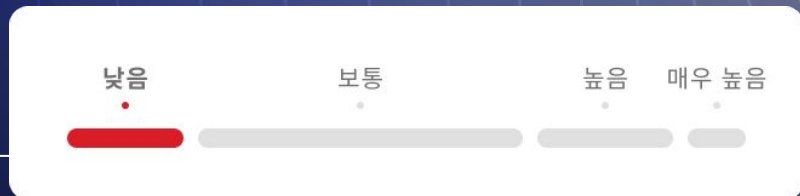
썸네일 주목도 '높음'

- 2021년~2023년 영상 중 조회수 **5,000** 이상
- 구독자수 / 조회수 **0.3~0.8** ('기타' 기준, 카테고리별 다름)
- 총 **68,073** 개 썸네일 이미지 (**23.8%**)



썸네일 주목도 '보통'

- 2021년~2023년 영상 중 조회수 **1,000** 이상
- 구독자수 / 조회수 **0.08~0.3** ('기타' 기준, 카테고리별 다름)
- 총 **86,076** 개 썸네일 이미지 (**30.1%**)



썸네일 주목도 낮음

- 2021년~2023년 영상 중 조회수 **100** 이상
- 구독자수 / 조회수 **0.08** 이하 ('기타' 기준, 카테고리별 다름)
- 총 **69,841** 개 썸네일 이미지 (**24.4%**)

데이터 전처리 - 리사이즈

.....



480 x 360

224 x 224

Splitfolder를 이용해서 train(0.6), test(0.2), validation(0.2) 나눔

모델 선정

Xception

val_accuracy: 0.2825

Resnet50

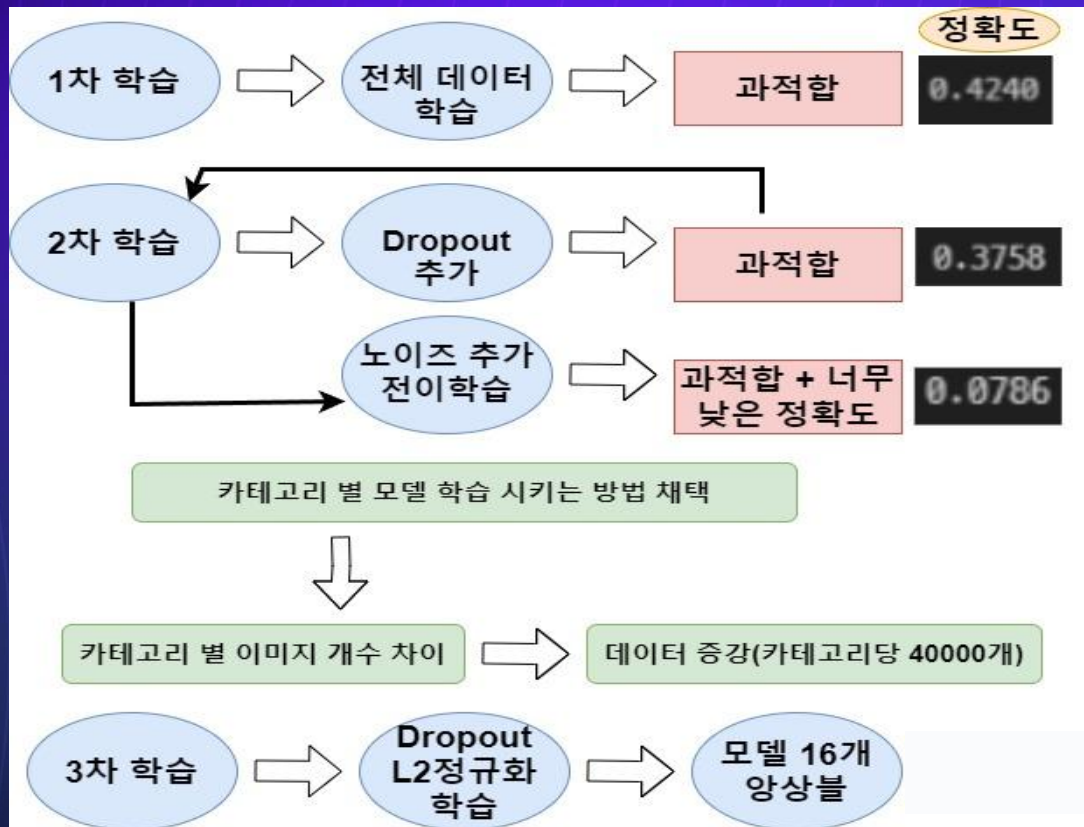
val_accuracy: 0.4064

InceptionV3

val_accuracy: 0.3631

정확도가 제일 높은 **Resnet50** 으로 모델 선정

Resnet50



한개의 이미지를
16개 모델의 예측확률을
평균해서 제일 높은확률 결정

학습된 모델 결과



Model 15 predicted class: 4
1/1 [=====] - 1s 641ms/step
Model 16 predicted class: 3
썸네일 주목도가 '매우 높음'입니다



1/1 [=====] - 1s 711ms/step
Model 16 predicted class: 3
썸네일 주목도가 '높음' 입니다



1/1 [=====] - 1s 887ms/step
Model 16 predicted class: 3
썸네일 주목도가 '낮음' 입니다

데이터/모델 관점 개선 사항



유튜브에서 사람들의 **Attention(관심)**을 잡아 끄는
썸네일 이미지를 찾기 위한 추가적인 고민들...



1. 데이터 라벨링 품질 -> 도메인 전문가가 썸네일에 라벨링 했다면?
2. 학습 데이터 양과 질 -> 데이터가 특정 채널, 카테고리, 스타일에 편중되지 않도록 고루 갖춰진다면?



1. 텍스트, 객체 인식 -> 이미지 안 텍스트와 객체가 영향이 어느 정도 있다면?
2. 기타 메타 데이터 -> 제목, 업로드 날짜, 채널, 태그 등 다른 메타 데이터를 학습에 활용할 수 있다면?

자체 평가 및 회고



1. 만족스러웠던 경험

- 다양한 이미지 분류 모델을 경험
- 실용적인 모델 방향성 결정
- 성능 개선을 위해 여러 방법을 시도해본것



2. 아쉬웠던 경험

- 한정된 일정 내 서비스 퀄리티의 모델 성능을 만들지 못했음
- 시간이 더 있었으면 더 성능을 올릴 수 있었을것 같아서 아쉽다



3. 이번 프로젝트를 통해 배운점

- 이미지 라벨링의 기준을 잡기가 어렵다
- 성능을 올리기 위한 전략을 과적합 해소 뿐만 아니라 다차원적으로 고민해볼 수 있었음

감사합니다