



YouHi

이태훈 이주형 김성수 김민재 이인평



INDEX

01. 피드백 반영

- 1). 차별화 전략 강조
- 2). 목표 개선
- 3). Object Detection 도입
- 4). 정확도 판단 기준

02. 진행 현황

- 1). 영상 검열
- 2). 음성 검열
- 3). Front 및 Back-end

03. 차 후 개발 계획



차별화 전략 강조

기존 시스템과의 차이점과
차별화 전략 강조.

다수 영상 플랫폼에는 존재하지 않는
자동 검열 시스템



Yellow dollar sign

정치적 발언, 선정적인 장면 테러 옹호 등이 포함된 영상에 광고가 달리는 것을
차단하기 위해 유튜브에서 시행하고 있는 광고주 친화정책

- 업로드 이후 이루어지는 검열방식
- 정확한 검열 사유를 업로더에게 알려주지 않는다.
- 영상 광고 수익 창출만을 제한.

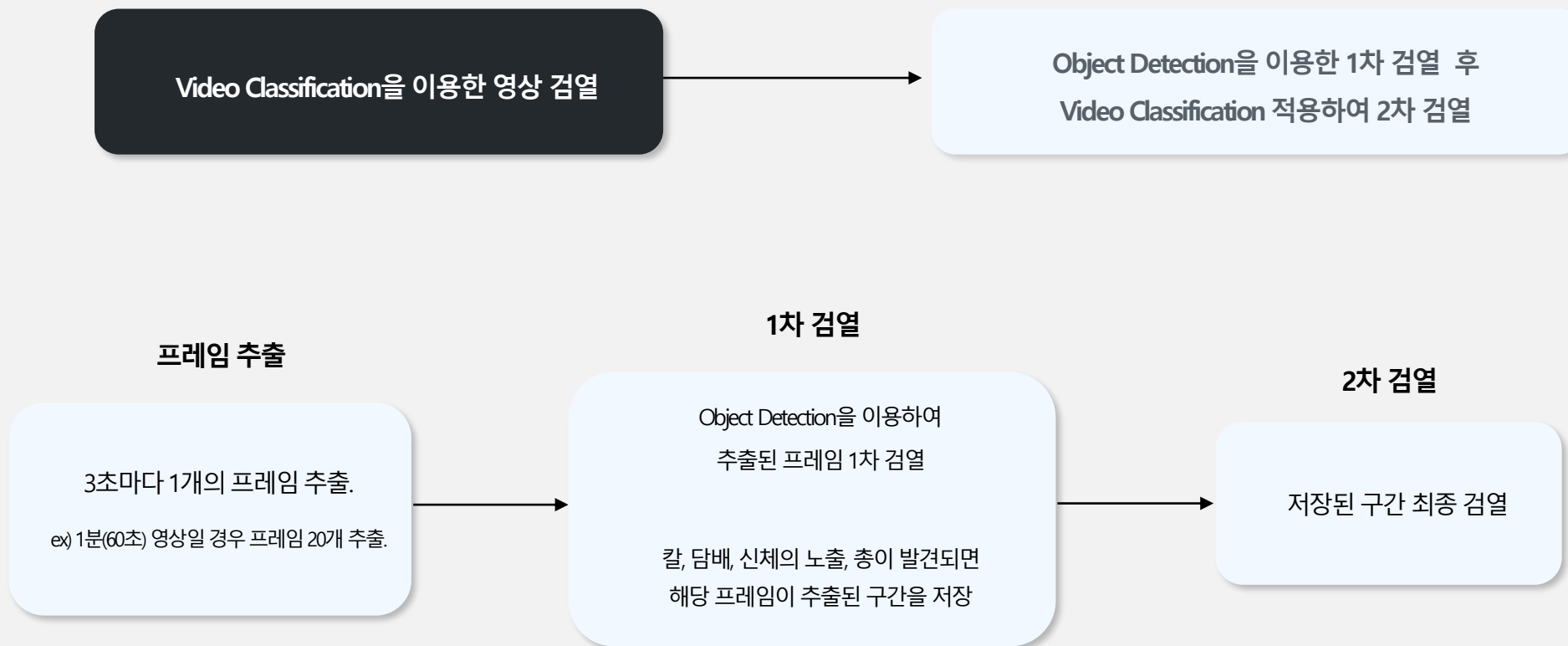


목표 개선





Object Detection 도입





정확도 판단 기준

검열의 정확도를 판단할 수 있는
기준이 필요하다

Precision / Recall 등 정량적인 수치를 이용하여
정확도 판단 기준 설정

		실제 정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

- True Positive(TP) : 실제 True인 정답을 True라고 예측 (정답)
- False Positive(FP) : 실제 False인 정답을 True라고 예측 (오답)
- False Negative(FN) : 실제 True인 정답을 False라고 예측 (오답)
- True Negative(TN) : 실제 False인 정답을 False라고 예측 (정답)

우리 프로젝트에 도입하게 되면 True는 유해한 것, False는 유해하지 않는 것

- **Precision(정밀도)** : 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율

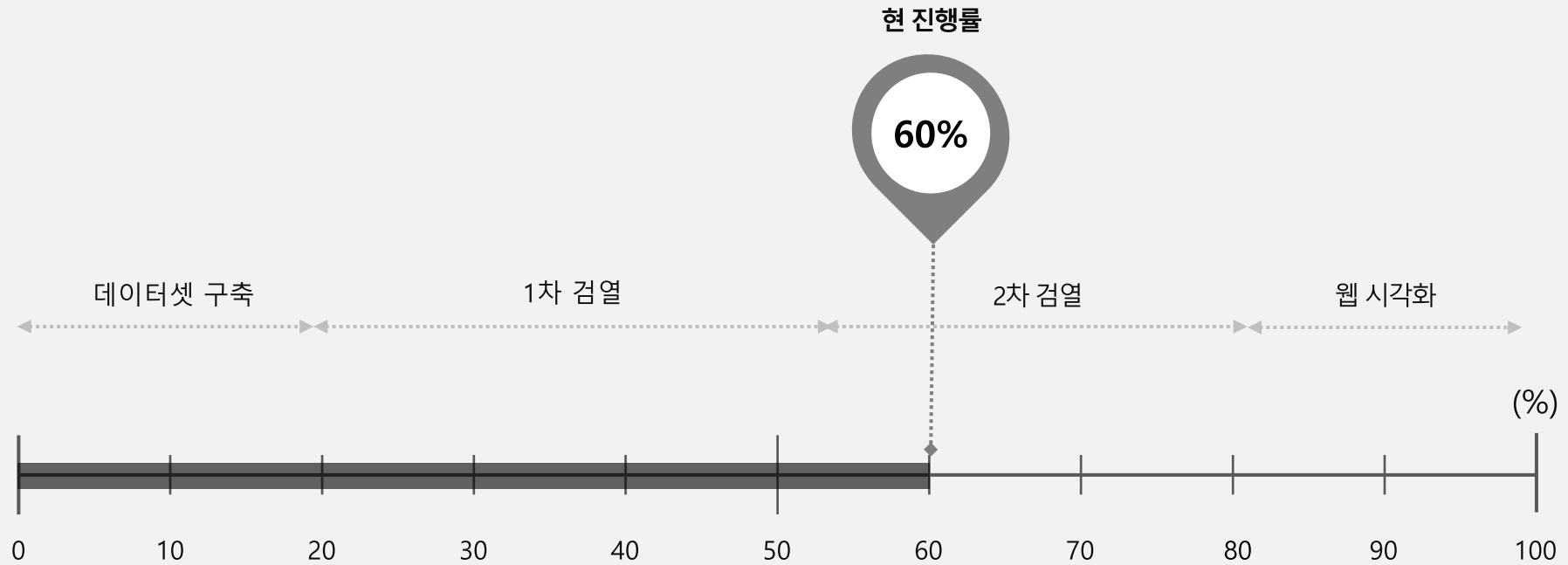
$$(Precision) = \frac{TP}{TP + FP}$$

- **Recall(재현율)** : 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율

$$(Recall) = \frac{TP}{TP + FN}$$



영상 검열 진행도





Object Detection - 데이터셋 구축

knife



smoke



game

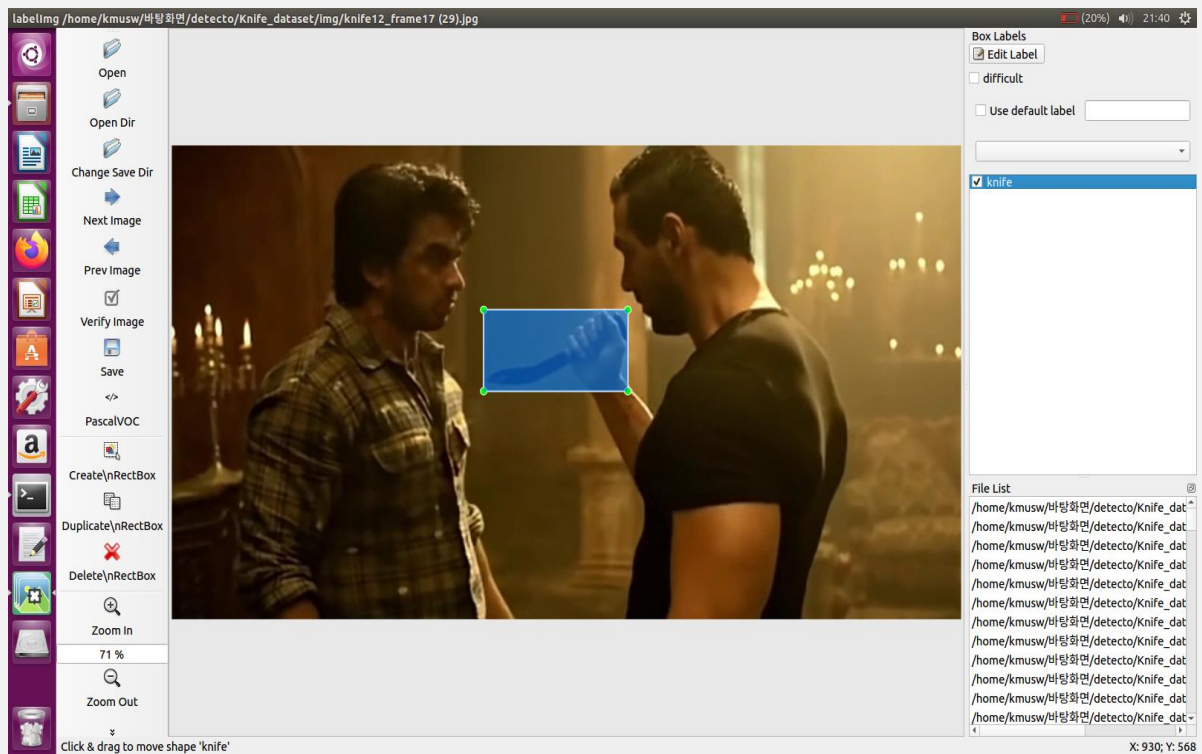


adult

...



200장씩



LabelImg를 이용한 라벨링 작업



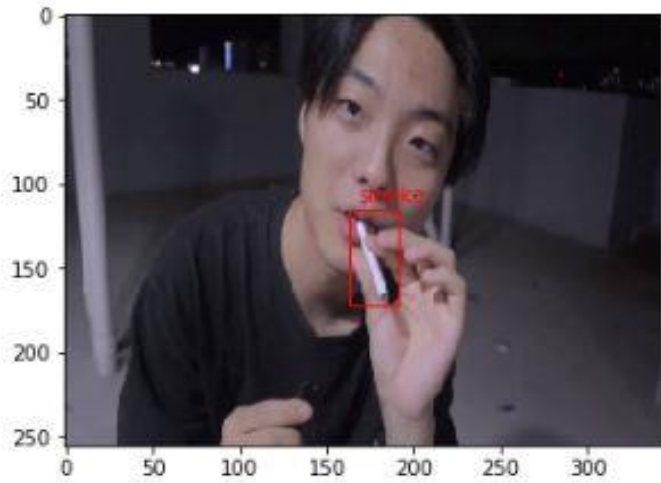
Object Detection - 결과

knife
tensor(0.6676)



Knife

['smoke']
tensor([0.9660])



Smoke

gun
tensor(0.8468)



Game



라벨에 따른 검열 방식

adult	game	smoke	knife
<p>Object Detection</p>	<p>Object Detection + Image Classification</p>	<p>Object Detection + Video Classification</p>	



Image Classification - game



Negative



Positive



Video Classification – knife, smoke

- Two Stream Convolutional Network

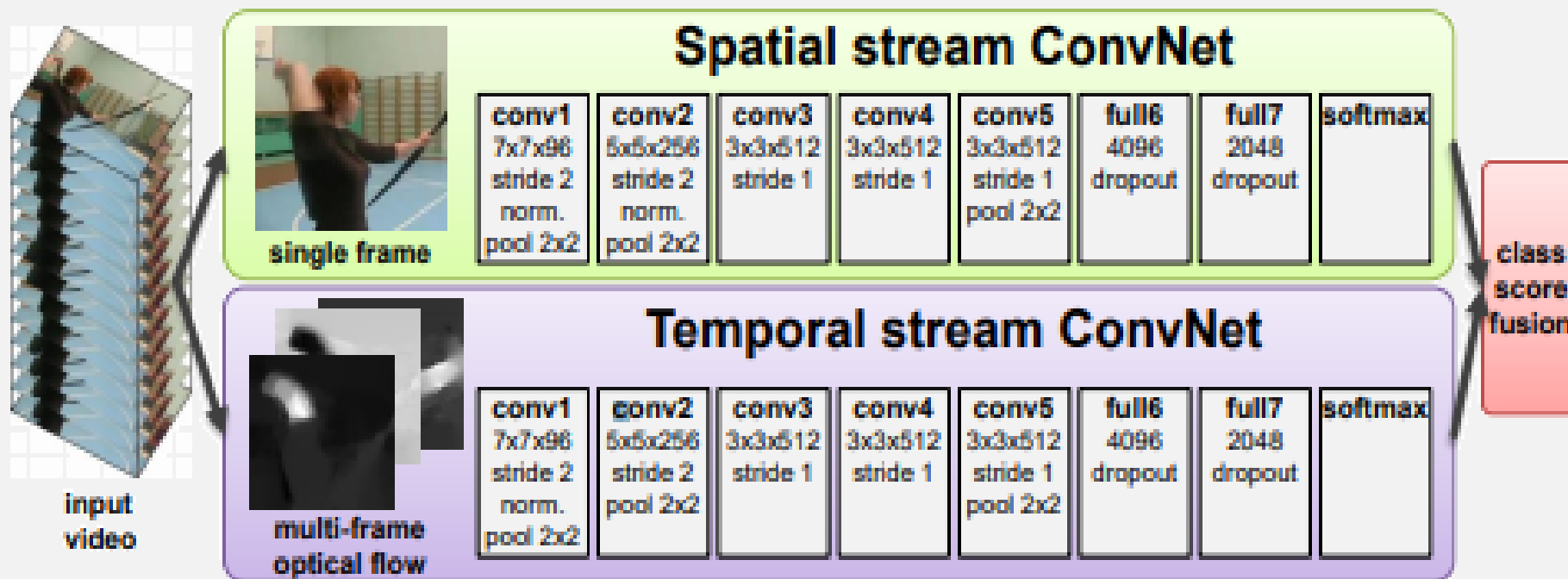


Figure 1: Two-stream architecture for video classification.



Video Classification – knife, smoke

- Two Stream Convolutional Network



Original image



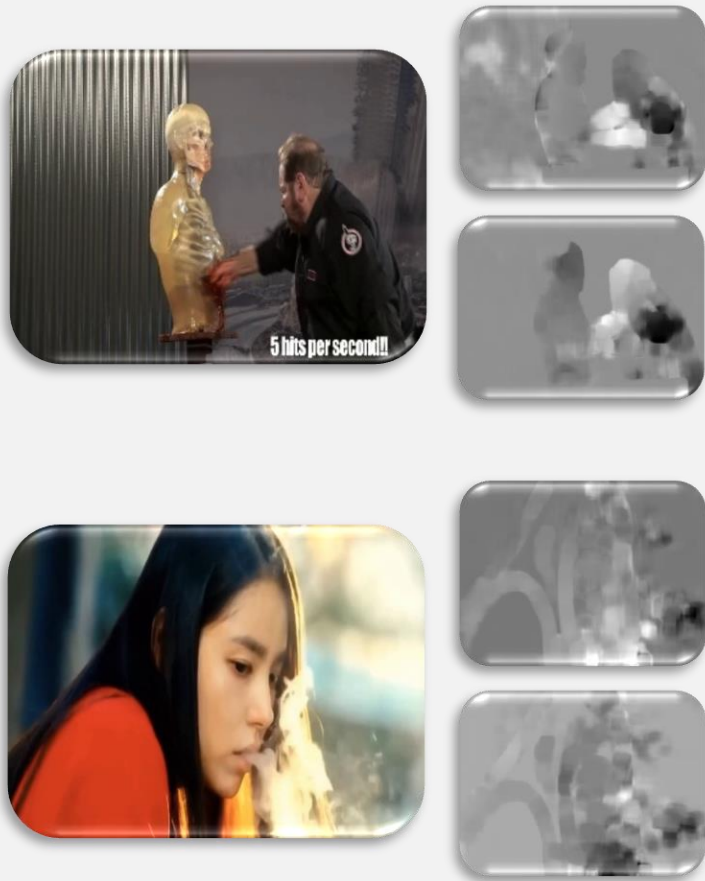
Horizontal optical flow



Vertical optical flow



Video Classification – knife, smoke



• • •



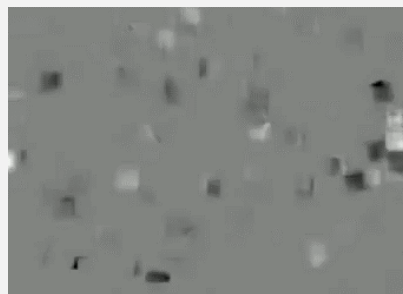
input video



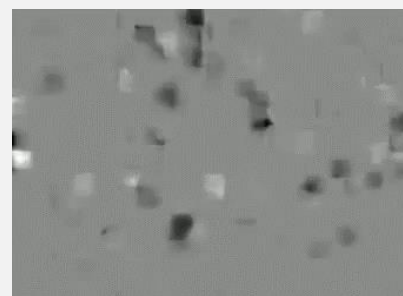
single frame



spatial stream



horizon flow



vertical flow



temporal stream

Class Score
Fusion

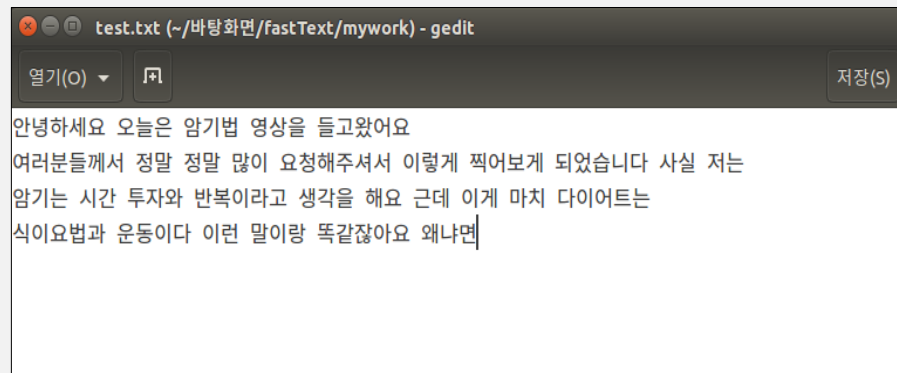
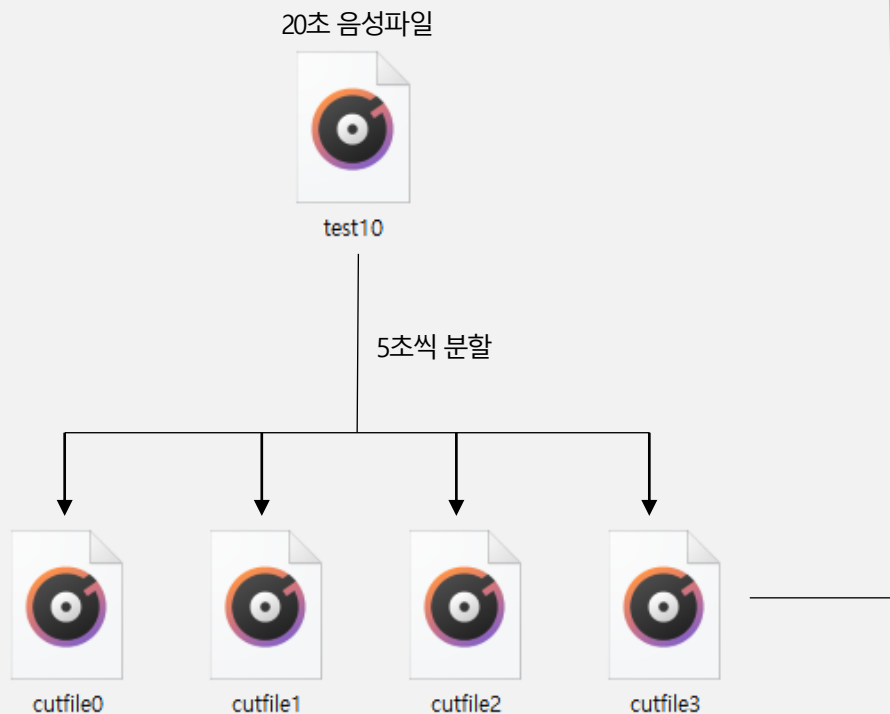


음성 검열 진행도

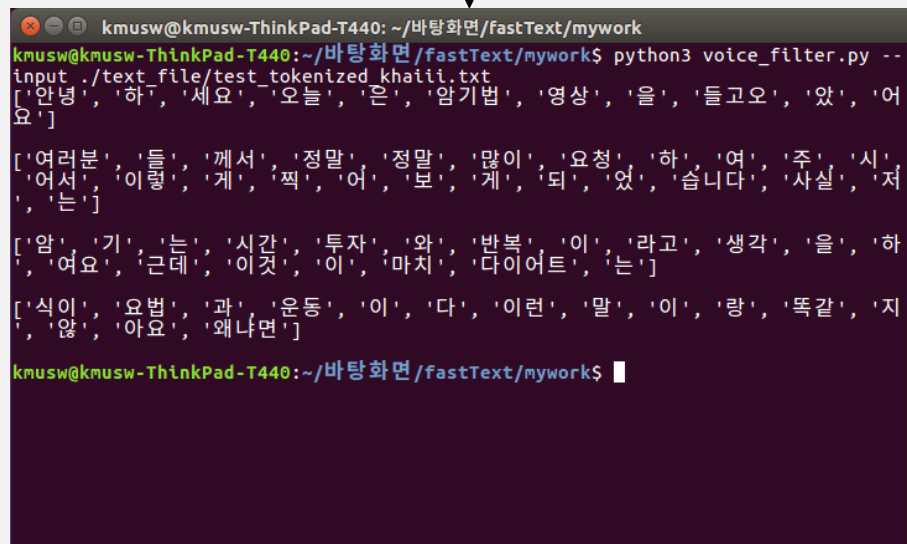




STT, 형태소 분석



line마다 음성파일 하나의 텍스트분량이 적혀있다.



Khaii를 이용하여 형태소를 분석한다.



진행 현황 - 음성 검열

FastText - 모델 학습

- 학습 데이터셋 = 한국어 위키피디아 텍스트 400MB + 추가적인 욕설 텍스트

```

kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork/text_file
무궁화대훈장(無窮花大勳章, Grand Order of Mugunghwa)은 대통령 및 그 배우자, 우방
원수 및 그 배우자 또는 대한민국의 발전과 안전보장에 기여한 공적이 뚜렷한 전직 우
원수 및 그 배우자에게 수여하며 등급은 없다.
건국훈장(建國勳章, Order of Merit for National Foundation)은 다음과 같다.
국민훈장(國民勳章, Order of Civil Merit)은 정치, 경제, 사회, 교육, 학술 분야에 공적
세워 국민의 복지 향상과 국가 발전에 기여한 공적이 뚜렷한 자에게 수여한다.
무공훈장(武功勳章, Order of Military Merit)은 전시 또는 이에 준하는 비상사태 하
에서 전투에 참가하여 뚜렷한 무공을 세운 자에게 수여한다.
근정훈장(勤政勳章, Order of Service Merit)은 공무원(군인 및 군무원을 제외한다)
및 사립학교의 교직원으로서 직무에 정려하여 공적이 뚜렷한 자에게 수여한다.
보국훈장(保國勳章, Order of National Security Merit)은 국가안전보장에 뚜렷한 공
을 세운 자에게 수여한다.
수교훈장(修交勳章, Order of Diplomatic Service Merit)은 국권의 신장 및 우방과의
친선에 공헌이 뚜렷한 자에게 수여한다. 수교훈장은 5등급으로 나뉘며 1등급 중 광화
대장은 외국의 수상급 이상, 광화장은 대사급 이상자에게 수여한다.
산업훈장(産業勳章, Order of Industrial Service Merit)은 국가산업발전에 기여한 공
적이 뚜렷한 자에게 수여한다.
새마을훈장(새마을勳章, Order of Saemaeul Service Merit)은 새마을운동을 통하여,
국가사회 발전에 기여한 공적이 뚜렷한 자에게 수여한다.
문화훈장(文化勳章, Order of Cultural Merit)은 문화·예술발전에 공을 세워 국민 문
화향상과 국가발전에 기여한 공적이 뚜렷한 자에게 수여한다.
체육훈장(體育勳章, Order of Sport Merit)은 체육발전에 공을 세워 국민체육 향상과
국가발전에 기여한 공적이 뚜렷한 자에게 수여한다.
과학기술훈장(科學技術勳章, Order of Science and Technological Merit)은 과학기술

```

```

kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork/text_file
고 년도에는 억 불 달성을 목표로 하고 있는 등 소비도시에서 생산도시로 변모하고 있
다. 특히 년 월까지 전국 최초로 개 권역에 문화산업 투자진흥지구 지정되고 년 월에
개 지구에 광주연구개발 특구가 지정되면서 미래성장동력 기반을 확충하였고 새로
일자리 만 개 창출과 자동차광가전그린에너지금형산업 등 대 주력산업을 중점 육성
하는 등 우리시의 소중한 자산이자 정적인 민주인권 문화예술 첨단산업 각축으로 한터
블러 사는 광주 더불어 행복한 시민도시 건설을 위해 광주가 새롭게 비상하고 있다.광
주광역시는 남쪽으로 남구 승촌동이 나주시 산포면과의 경계를 이루고 북쪽으로는 북
구 우치동이 동쪽으로 남구 용강동 영산강변 동부지역이 화순군 남면과의 경계를 이루고
있다 서쪽은 광산구 양동이 함평군 나산면과 경계를 이루고 있다 동서간의 남해도시로
장거리 간격은 각각 이다 대한민국 서남부에 위치한 광주는 호남 지방의 중핵도시로
서 역할을 수행했다 광주를 기점으로 하여 나주시 무안군 목포시를 잇는 광무선 영암군 순
창군 완도군 완도의 광안선 장성군 정읍시 전주시 여수시를 연결하는 광전선 그리고 광안
부안군을 연결하는 광부선 광부선 등 개 방향의 중심축으로서 정치 경제 사회 문화 교육
제반에 걸쳐 광주는 동부 산악지대와 서부 평야지대의 접이 지대라고 할 수 있다 광주의 산
하고 형태는 노년기에 속하여 대체로 준원화 된 구릉성 지대라고 보 된다 광주의 중심인
한 해는 전남 지방의 산세를 보며 북쪽에 노령산맥이 동서방향으로 뻗어 전복과 자연적인
경계를 이루고 동쪽의 산맥을 소백산맥과 그 기맥들이 남북방향으로 뻗어 경남과 경서를 이
루며 전남의 남쪽의 산맥이 울타리 사이에 자리한다 광주평야는 노령산맥과 소백산맥의 지
로 맥들 사이의 고령산맥의 양쪽에 대상을 포하며 광주천과 그 소지류 및 극락강 수역에 연
맥을 나타내고 있다 광주평야는 대상을 따라 인접한 나주 등지로 이어져 결국 나주평야의
인부현에 가까운 특성을 보이고 있다

```

```

kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork
Progress: 2.9% words/sec/thread: 11405 lr: 0.048567 avg.loss: 1.857290 ETA
Progress: 2.9% words/sec/thread: 11404 lr: 0.048564 avg.loss: 1.856907 ETA
Progress: 2.9% words/sec/thread: 11407 lr: 0.048560 avg.loss: 1.856780 ETA
Progress: 2.9% words/sec/thread: 11408 lr: 0.048557 avg.loss: 1.856559 ETA
Progress: 2.9% words/sec/thread: 11411 lr: 0.048554 avg.loss: 1.856321 ETA
Progress: 2.9% words/sec/thread: 11414 lr: 0.048550 avg.loss: 1.855682 ETA
Progress: 2.9% words/sec/thread: 11415 lr: 0.048547 avg.loss: 1.855424 ETA
Progress: 2.9% words/sec/thread: 11416 lr: 0.048543 avg.loss: 1.855334 ETA
Progress: 2.9% words/sec/thread: 11420 lr: 0.048540 avg.loss: 1.855106 ETA
Progress: 2.9% words/sec/thread: 11421 lr: 0.048536 avg.loss: 1.854860 ETA
Progress: 2.9% words/sec/thread: 11422 lr: 0.048533 avg.loss: 1.854154 ETA
Progress: 2.9% words/sec/thread: 11425 lr: 0.048529 avg.loss: 1.853499 ETA
Progress: 2.9% words/sec/thread: 11424 lr: 0.048526 avg.loss: 1.853492 ETA
Progress: 3.0% words/sec/thread: 11427 lr: 0.048523 avg.loss: 1.853018 ETA
Progress: 3.0% words/sec/thread: 11426 lr: 0.048520 avg.loss: 1.852681 ETA
Progress: 3.0% words/sec/thread: 11427 lr: 0.048516 avg.loss: 1.852060 ETA
Progress: 3.0% words/sec/thread: 11427 lr: 0.048513 avg.loss: 1.851483 ETA
Progress: 3.0% words/sec/thread: 11428 lr: 0.048510 avg.loss: 1.851539 ETA
Progress: 3.0% words/sec/thread: 11427 lr: 0.048507 avg.loss: 1.851097 ETA
Progress: 3.0% words/sec/thread: 11430 lr: 0.048503 avg.loss: 1.850826 ETA
Progress: 3.0% words/sec/thread: 11429 lr: 0.048500 avg.loss: 1.850845 ETA
Progress: 3.0% words/sec/thread: 11432 lr: 0.048496 avg.loss: 1.850558 ETA
Progress: 3.0% words/sec/thread: 11435 lr: 0.048493 avg.loss: 1.850266 ETA
: 0h25m14s

```

- 전처리되어진 텍스트 데이터셋으로 FastText 모델 학습.



FastText – 유사도 검사

- 학습된 FastText 모델을 이용하여 전처리된 형태소와 욕설의 cosine similarity 계산

```
kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork
kmusw@kmusw-ThinkPad-T440:~/바탕화면/fastText/mywork$ python3 voice_filter.py --input ./text_file/test_tokenized_khaii.txt
['안녕', '하', '세요', '오늘', '은', '암기법', '영상', '을', '들고오', '왔', '여요']

['여러분', '들', '께서', '정말', '정말', '많이', '요청', '하', '여', '주', '시', '어서', '이런', '게', '찍', '어', '보', '게', '되', '었', '습니다', '사실', '저', '는']

['암', '기', '는', '시간', '투자', '와', '반복', '이', '라고', '생각', '을', '하', '여요', '근데', '이것', '이', '마치', '다이어트', '는']

['식이', '요법', '과', '운동', '이', '다', '이런', '말', '이', '랑', '똑같', '지', '않', '아요', '왜냐면']

kmusw@kmusw-ThinkPad-T440:~/바탕화면/fastText/mywork$
```

- Khایی를 이용한 형태소 분석 결과

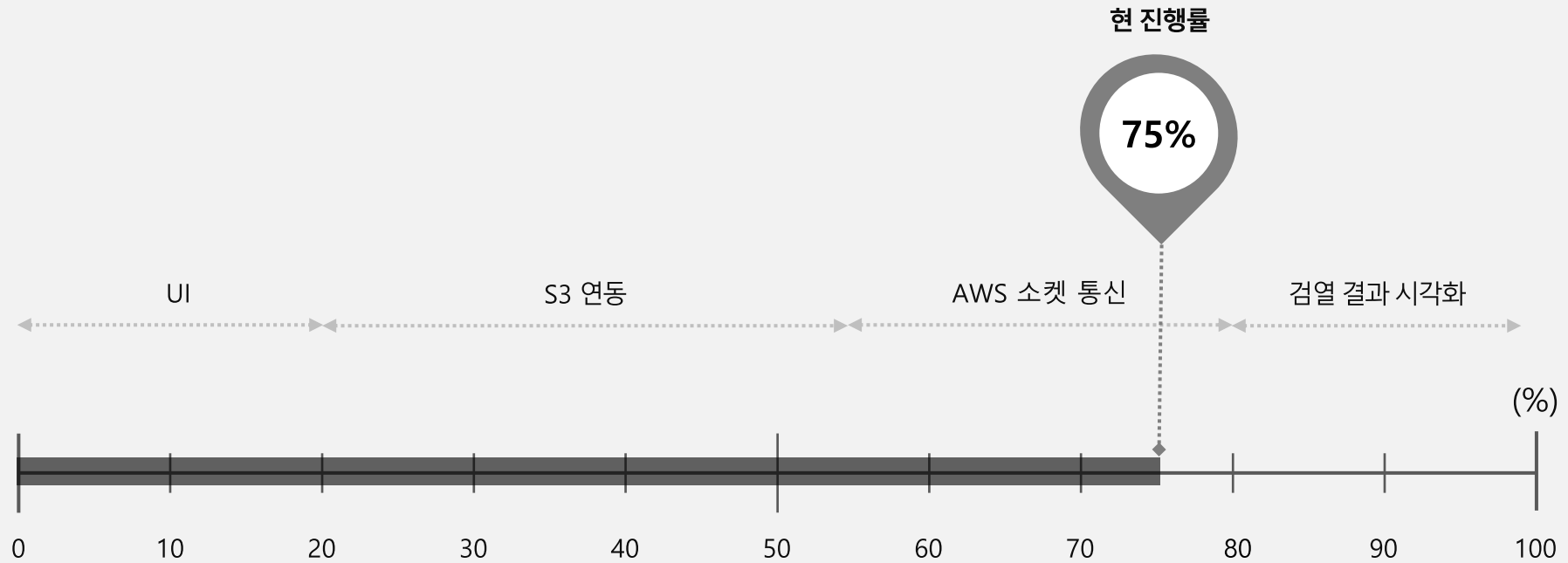
```
kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork
암: 0.12583464
기: 0.11561203
는: 0.13864249
시간: 0.3251203
투자: 0.17260684
와: 0.16482943
반복: 0.36835143
이: 0.20231715
라고: 0.37160045
생각: 0.5230171
을: 0.18702115
하: 0.54145336
여요: 0.18268977
근데: 0.8071771
이것: 0.32653284
이: 0.20231715
마치: 0.37377688
다이어트: 0.2923674
```

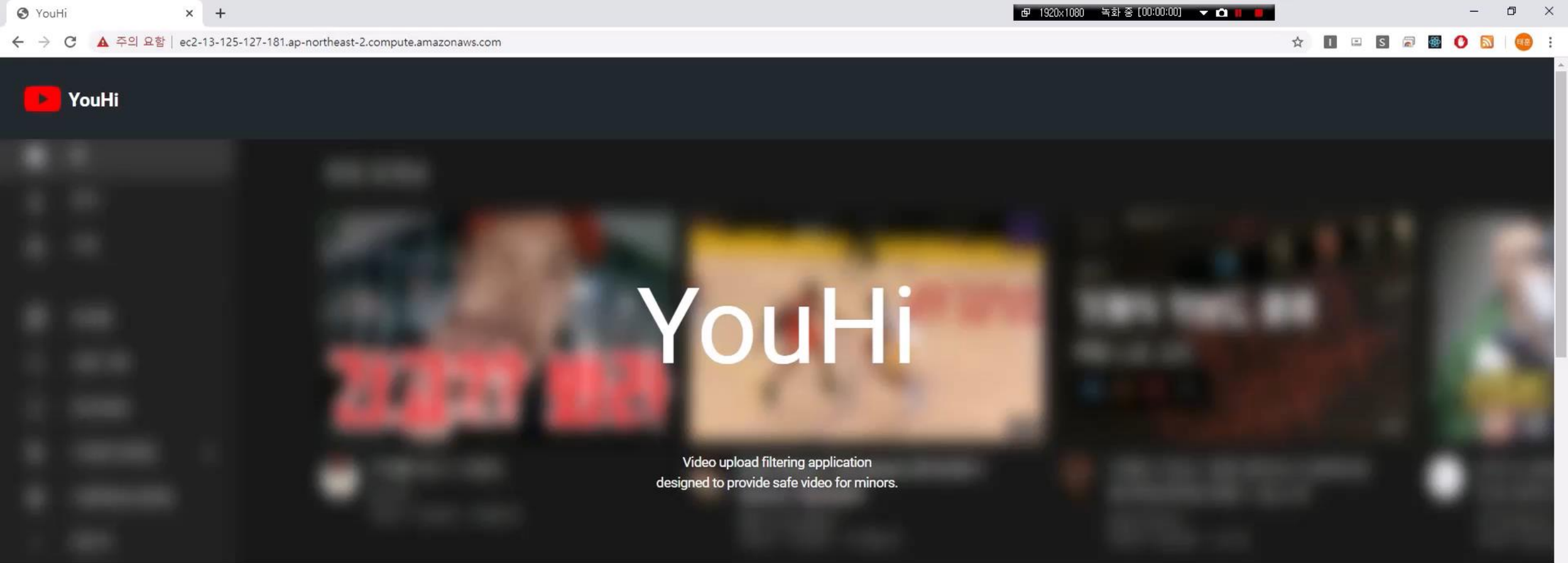
차 후 String Matching을 통해 검열 예정

- 욕설 '시X' 와 각 형태소 간의 cosine similarity 결과



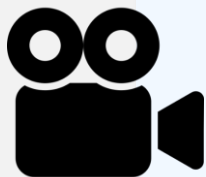
Front 및 Back-end 진행도





YouHi is

YouHi는 영상분류 신경망 모델을 이용한 정확한 필터링을 제공함으로써
업로더의 동영상에서 청소년에게 부적절한 내용을 모두 걸러냅니다.
필터링의 기준은 YouTube 가이드라인의 연령 제한 콘텐츠에 명시되어 있는
연령 제한 적용 고려 사항이며 자세한 내용은 여기를 참고하세요.



영상 검열

- 2차 검열



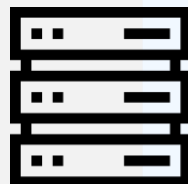
Front-end

- 검열 결과 시각화



음성 검열

- 스트링 매칭



Back-end

- 유지 보수



Thank you.

들어주셔서 감사합니다.