Thirty, thirsty for love

- 서른, 사랑에 목마르다

허유나, 최성진, 이태형, 배보람

Contents . . . ♥

주제 선정 배경 . . . ♥

요즘 2030세대 연애도 안해..."미혼자의 21.1%, 연애 경험 없는 '모솔'" (영상)

요즘 2030 세대에서 '비연애주의'가 급증하면서 결혼율 및 저출산 문제가 더욱 심각해지고 있다.

최민서 기자 입력 2023.03.04 15:36

기사의 이해를 돕기 위한 자료 사진 / gettyimagesBank

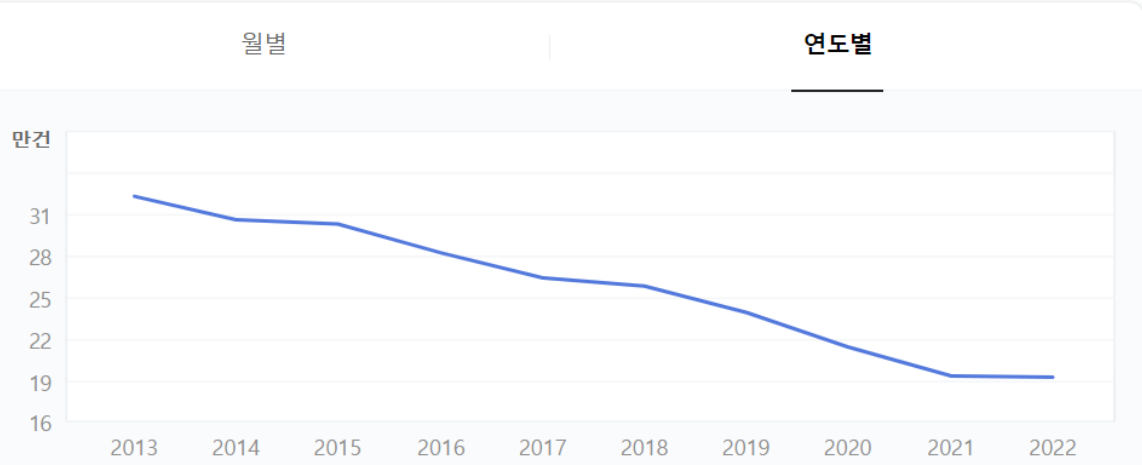
누적판매
초대용량
50% + 브
메디큐브

베스트클
건강
"고기 많
나"...'AD
커진다

혼인건수

월간 **1만 6,053건** '23.06

연간 **19만 1,690건** '22



출처 KOSIS

통계청 KOSIS 지표 더보기 →

출처 : 인사이트 뉴스 https://www.insight.co.kr/news/431457

나는 SOLO

주제 선정 배경 . . . ♥

사회적 문제 + 화제성 + 대중의 흥미

나는 Iam SOLO

설 명 . . . ♥

'나는 솔로'

- 결혼을 간절히 원하는 12명의 솔로 남녀들이 모여 사랑을 찾기 위해 고군분투하는 극사실주의 데이팅 프로그램
- 각 출연자는 외모와 분위기에 따라 제작진이 정해준 '이름'을 받음

내가 '나는 솔로'에 출연한다면?(가제)

사용자의 사진을 넣으면 그 사진의 이미지을 분석하여 어울리는 '이름'을 보여주고

'이름'이 어떤 사람인지 글과 사진을 통해 설명하는 페이지를 제작

나는 SOLO

요구사항 정의서 . . . ♥

요구사항 정의서

개발환경	o/s : window	언어	python			
대분류	중분류	소분류	상세설명	우선순위	담당자	라이브러리
이미지 처리	1. 데이터 수집 및 정제	1-1. 이미지 탐색 및 저장	Fatkun Batch 활용 : Naver, Google 이미지 분류 및 저장	1	배보람, 최성진, 허유나	
		1-2. 이미지 비율 설정	340*160으로 이미지 비율 설정	1	배보람	matplotlib, cv2, os, glob
		1-3. 배경 이미지 삭제	배경이미지 삭제	1	배보람	
이미지 모델링	2. 데이터 전처리	2-1. 이미지 라벨링	나는 솔로 출연자 이름별 라벨링 진행	2	이태형, 최성진	rembg, pandas, numpy
		2-2. 이미지 증폭	albumentations 회전, 반전, 노이즈, 밝기대비를 활용한 증폭	2	이태형, 최성진	joblib, albumentations
		2-3. 이미지 데이터화	이미지데이터 이름별로 데이터프레임화 후 병합	2	이태형, 최성진	
	3. 모델링	3-1. 이미지 데이터 학습 및 검증	RandomForest Regression 모델 구축	3	이태형	sklearn, autogluon
서버구축	4. 서버구축	4-1. HTML, python 서버 연동	웹 인코딩 후 학습데이터 읽기	4	이태형	cgi, sys, codecs
		4-2. 입력 이미지 저장	웹 페이지 Form -> Input 이미지 저장	4	이태형	
		4-3. 예측 모델 적용	예측 모델로 사진 판정 후 출력	4	이태형	
HTML	5. 웹페이지 구성	5-1. 메인페이지	메인페이지 디자인 설계	5	이태형	
		5-2. 서브페이지	서브페이지 디자인 설계	5	배보람, 허유나	
		5-3 서브페이지 자료조사	이미지 수집	5	배보람, 허유나	
		5-4 서브페이지 내용 구성	자료 수집 및 내용 편집	6	배보람, 허유나	
PPT	6. PPT 제작	6-1 PPT 디자인 및 구성	PPT 제작	6	배보람	

나는 Iam SOLO

데이터 전처리 . . . ♥

- 출연자 12명 각각 이미지 100~300장 수집
- 이미지 파일의 grayscale작업 / 배경 제거/340*160
- nd.array로 바꾸어 npz로 저장하는 함수 생성
- 라이브러리(Albumentations)를 사용하여 이미지 증강
- 총 20만여장의 이미지 사용
- 각 성별당 6개의 이름 카테고리로 라벨링
- 증강한 이미지를 성별로 나누어 DataFrame으로 병합

나는 Iam SOLO

모델링 . . . ♥

- LightGBM, CatBoost, XGBoost를 처음에 고려 했지만 부스팅 방식이라 모델 학습 시간을 계산 해 보았을 때 3일이 넘게 걸려서 그나마 가장 성능이 좋은 RandomForest로 모델을 선정

- 학습 중 ram 부족으로 에러가 나서 int64를 int16으로 수정

- Train : 97, test : 70이지만 여러 사람을 하나의 이름으로 학습 시킨 것이라 좋은 성능이라고 판단

```
================test================
0.6955390334572491
              precision    recall  f1-score   support

           1       0.82      0.55      0.66       382
           2       0.73      0.69      0.71       471
           3       0.68      0.81      0.74       493
           4       0.66      0.75      0.70       474
           5       0.68      0.64      0.66       438
           6       0.68      0.70      0.69       432

    accuracy                           0.70      2690
   macro avg       0.71      0.69      0.69      2690
weighted avg       0.70      0.70      0.69      2690

================train================
0.9704351059873559
              precision    recall  f1-score   support

           1       0.99      0.96      0.98      1530
           2       0.97      0.97      0.97      1881
           3       0.96      0.98      0.97      1971
           4       0.96      0.97      0.97      1894
           5       0.97      0.96      0.97      1752
           6       0.97      0.97      0.97      1728

    accuracy                           0.97     10756
   macro avg       0.97      0.97      0.97     10756
weighted avg       0.97      0.97      0.97     10756
```

나는 SOLO