

# 37 뉴스 중복 필터링 애플리케이션

소속 정보컴퓨터공학부

분과 C

팀명 JKO

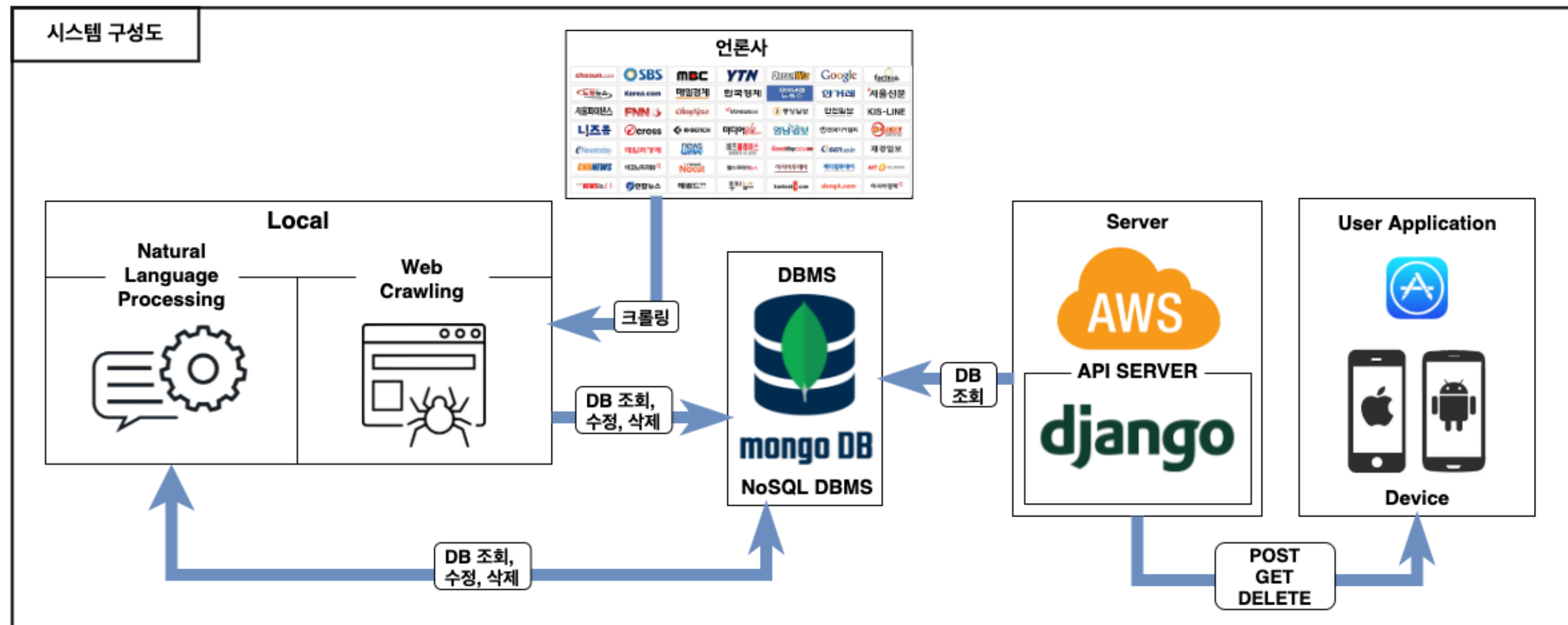
참여학생 김동주, 오상현, 조현빈

지도교수 유명환

## 작품 구성 및 상세 내용

- ❖ 유사 주제를 갖는 대량의 뉴스 기사들을 하나의 뉴스 기사로 제공하는 애플리케이션을 개발한다.
- ❖ 사람의 감정을 기준으로 뉴스 기사를 분류하는 모델을 개발한다.

## 작품 구성 및 상세 내용



- 1) 언론사 별 뉴스들을 크롤링 후, DB에 저장한다.
- 2) 유사도 측정을 통해 뉴스들을 분류한다.
- 3) 감정 분류 모델을 통해 뉴스들을 분류한다.
- 4) API 서버에 전송한다.
- 5) 모바일 애플리케이션을 통해 사용자에게 제공한다.

## 감정 분석 과정 및 결과

### 데이터 셋 준비

|          |                                                 |   |
|----------|-------------------------------------------------|---|
| sport    | '별들의 전쟁' PGA 투어 플레이오프 시작                        | 2 |
| economy  | '코로나나 대출' 안기는 계속 연장... 또 다른 '열대 폭탄' 키우나         | 0 |
| sport    | '소속 팀 부진' 동료와 불화설' 조남두, 유원투스에 날을 전함             | 0 |
| economy  | '코로나19 자영업 직격탄' 속 역대 매출 '눈에 띄네'~                | 1 |
| sport    | '손흥민 72분' 트로트, 올버튼 1-0 격고 개막 2연승                | 1 |
| politics | "언론재갈법 진자 육작은 외 집권연장"... 총력저지 외전 쉼              | 0 |
| sport    | '아시아 1인자' 패럴림픽 초청 김세정 "기량 110% 만들 것"            | 2 |
| sport    | '달타이 역전포 NC, 3연승 일주... 키움, 3연패 활물               | 1 |
| society  | 부산 코르나19 118명 추가 확진... 육육향, 체육시설 등 연쇄감염 확산      | 2 |
| economy  | 'LPGA 마지막 메이저' AIG 여자오픈, 홍상근 68홀원... 역대 최고역     | 1 |
| economy  | 'VR로 집 보고, AI로 집짓는다'... 일상에 과고든 '프롤트르' [부동산360] | 2 |
| society  | 부산 중구 의류급여대상자 사례관리 '킬리망자로 마을에 색을 입히다'           | 4 |
| economy  | "2030년까지 재생에너지 비중 20%로 늘리기 어렵다"                 | 2 |
| politics | 결국 나선 유승민 '윤석열, '이준석 흔들기' 직접 사과하라"              | 0 |
| politics | 문대통령, 26일 'K+ 벤처' 행사... "글로벌 벤처강국 도약"           | 1 |
| politics | 언론중재법' 김의겸 손에 달린다... 외, 안전조정의 표결 처리 수는 돌입       | 0 |
| society  | 부산 중구 중앙동 지역사회보장협의체, '이웃사랑' 등복! 친환경로마포' 전달      | 0 |
| politics | 속근 22사단장, 부임 5개월 만 해임... 성주영 2차 가해 의혹           | 0 |
| society  | 부산 황정산에 해발 500m 전망대... 대규모 유원지조 개발              | 4 |
| society  | 부산 NH농협은행 대원동지점, 농림 최고 영예 '문화상' 수상              | 1 |

총 뉴스의 4개면 정치, 사회, 경제, 스포츠 (약 5000개)

<라벨링 정보>

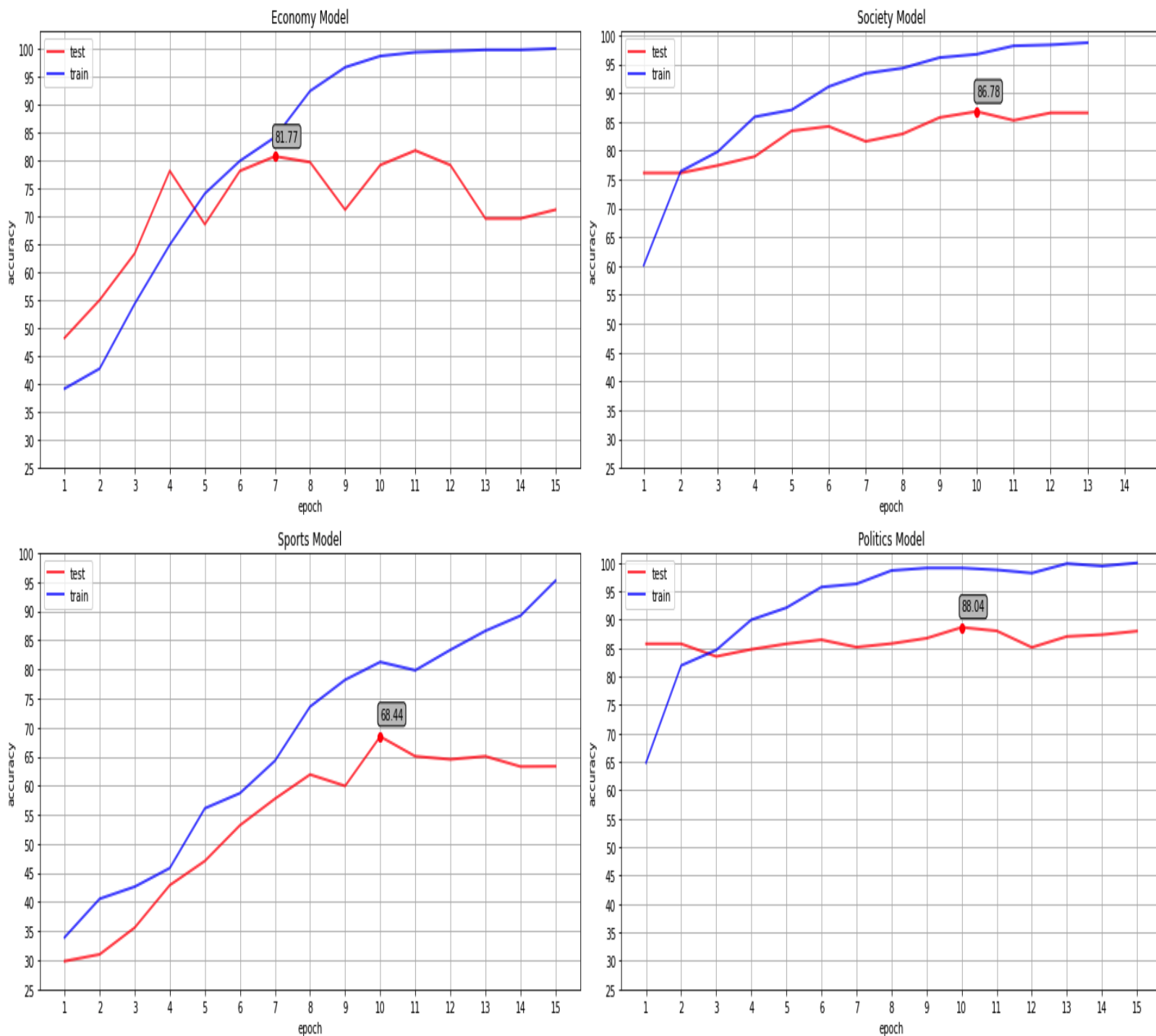
정치 : 부정(0)/긍정(1)

사회 : 따뜻한(0)/신기한(1)/충격적인(2)/슬픈(3)/중립(4)

경제 : 부정(0)/긍정(1)/중립(2)

스포츠 : 부정(0)/긍정(1)/중립(2)

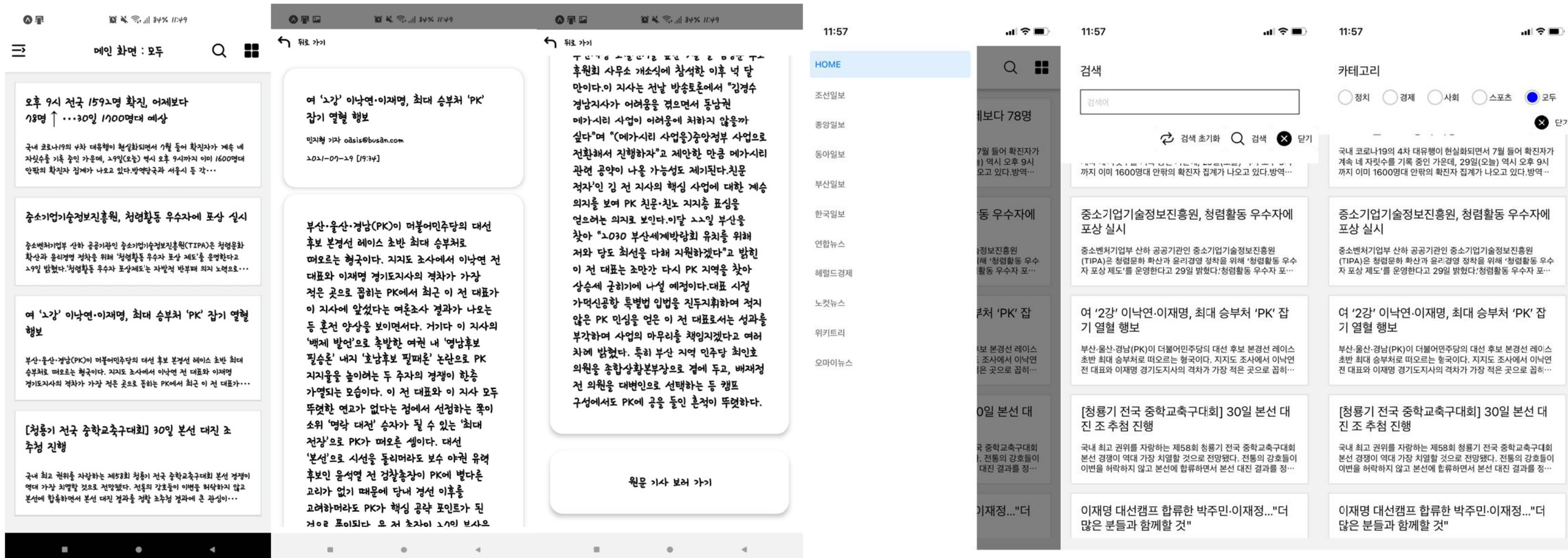
### 모델 학습 및 평가



### 분석 결과

|       |                                                                                     |
|-------|-------------------------------------------------------------------------------------|
| ❖ 정치  | [13] predict('서북 해군 부산상 성추행 사망 2차 가해 '날남지' 수사할 것', model_loaded)<br>부정<br>'0'       |
|       | [14] predict('군 성범죄, 민간 병원서 재판...법사위 소위 통과', model_loaded)<br>긍정<br>'1'             |
| ❖ 사회  | [20] predict('양행군 김장태소 토종씨앗 나눔 행사 개최', model_loaded)<br>따뜻한<br>'0'                  |
|       | [21] predict('박이정' 먹는 물로 물산자기부, 대양' 열었던 담수화 장치 개발', model_loaded)<br>신기한<br>'1'     |
|       | [22] predict('죽어갔다' 배달 중 다른 배달원과 사비 불자 불고 함박한 20대 실행', model_loaded)<br>충격적인<br>'2' |
|       | [23] predict('금강하곡동 통신문 지니면 아선 전복 1명 사망 5명 구조', model_loaded)<br>슬픈<br>'3'          |
| ❖ 경제  | [ ] predict('공정위 미래에너지 개발사 불입 대출 혐의 조사 착수', model_loaded)<br>부정<br>'0'              |
|       | [ ] predict('상반기 카드 많이 끊었다 카드사 손이익 33.7% 증가', model_loaded)<br>긍정<br>'1'            |
| ❖ 스포츠 | predict('불쌍한 버려진 바르셀로나 이레 스텔라 스텔라인도 일제히 포기', model_loaded)<br>부정<br>'0'             |
|       | predict('패럴림픽 육상 100m 세계신기록 달성', model_loaded)<br>긍정<br>'1'                         |

## 결과 및 응용



- ❖ 뉴스를 구독하는 사용자의 편의에 맞춰 중복된 뉴스를 묶어 대표 뉴스 하나만 보여줌으로써 사용자의 편의성을 높였다.

- ❖ 뉴스의 기사 제목을 감정 분석하여 분류함으로써 사용자가 뉴스를 선택하는 폭을 넓일 수 있도록 했다.