

# 딥러닝 기반 핵심 산업별 빅데이터 분석

## <키워드 분석 프로젝트 - 5팀>

일 자	2018년 11월 19일	주 제	정치 키워드 분석
팀 장	정민호<swagmh@naver.com>	팀 원	유선우, 조민정

### 주제 | 2018년 문재인정부 정책 중간 평가

2018년 언론사와 국민청원 게시판의 기사와 글들을 통해 2018년 문재인 정부 정책에 대하여 점수평가 하기

#### 1. 정책 TOP 3 뽑기

##### 1-1. 데이터 크롤링

수집데이터1. 언론(보수/중도/진보) 선정 - 조선/세계/한겨레  
각 언론사별 정치, 경제, 사회 기사 타이틀 크롤링

##### 1-2. 정부 100대 과제와 비교하여 대주제 선정

1-1에서 나온 데이터를 분석하여 정치, 사회, 경제 각 파트에서 가장 연관성 높은 정책 선정

#### 2. TOP3 정책 관련 연관어 분석

2-1. 각 분야별 워드클라우드 분석을 통하여 상위 빈도 단어 도출.

2-2. 상위 빈도수에서 단어를 주제별로 분류하여 연관 이미지 재분석.

#### 3. 국민청원 및 네이버 데이터랩을 통해 주제별 긍/부정 이미지 분석

3-1. 국민청원 관련 게시판 별 데이터 크롤링 및 워드클라우드 분석

3-2. 주제에 대한 단어 도출

3-3. 3-2.이미지를 통해 네이버 데이터랩에서 검색 데이터 수집

3-4. 3-3 데이터를 통해 정책에 대한 긍/부정 이미지 도출

## 4. 점수화 과정

4-1. 점수 지표 기준: 청원점수 25%, 네이버랩 검색지수 25%, 목표달성도 50%

- 주제가 정책을 평가하는 것이기 때문에 실질적인 이행률의 비율을 가장 높게 두었고, 여론 반응의 경우에는 부가적인 것이라고 생각해서 청원과 검색지수의 점수를 나누었음. 또한 청원의 경우, 대부분 정책에 대한 불만을 표시하는 것이기 때문에 모두 마이너스 점수로 처리하였음. 네이버랩 검색지수는 긍정, 부정 이미지에 대한 관심도를 나타내는 지수로 유의미함.

4-2. 정책평가

- 2018년 기획재정부 정책 예산안에 기재된 예상 정책을 기준으로 실제 이행률과 실효성을 판단하여 점수화 하였음. 정량적인 수치 변화에 정책 실효성(예산 사용은 했으나 내역이 불투명한 경우 등)을 종합적으로 판단하여 점수를 매겼음.

**감사합니다**