

# M'AI Report (마이 리포트)

SNS & 언론매체 분석을 통한 생성형 AI 리포트  
Media-based Generative Public Policy & Legal Analysis Report

D.P.

01

프로젝트 개요

제안 배경  
프로젝트 소개  
목표 결과물

02

활용 데이터

데이터 개요

03

모델 개발 방법

Project Architecture  
서비스별 적용 기술

04

실험 및 평가

데이터 전처리  
EDA 결과 보고  
상세 내용

05

활용 계획  
및 기대효과

시각화 개요  
시각화 목적  
결과물 소개  
기대효과

06

시연

모델 시연  
Tableau 시연

01

# Outline

**프로젝트 개요**

## 1-1 제안 배경

### 방대한 뉴스와 SNS 데이터 속, 대중의 의견을 반영하는 여론 분석 모델의 부재



#### Problem

##### 1 정책과 여론의 거리감

낮은 정책 체감도, 사각지대 지속 발생

##### 2 방대한 여론 데이터 분석 한계

시간/비용 문제 및 데이터 해석의 어려움

##### 3 전담 연구기관 부재\*

정책 발굴, 실태 통계 관리 등 근거 기반 정책 수행에 한계

\*제9차 청년정책조정위원회

#### Solution

##### 1 여론-정책 연계성 강화

시간 흐름에 따른 여론 분석, 지역별 분석을 통한 여론 적시 탐지 및 사각지대 해소

##### 2 AI 분석 모델을 통한 정책 방향성 제안

분석 시간과 비용 감소 기여 및 핵심 주제 해석 용이

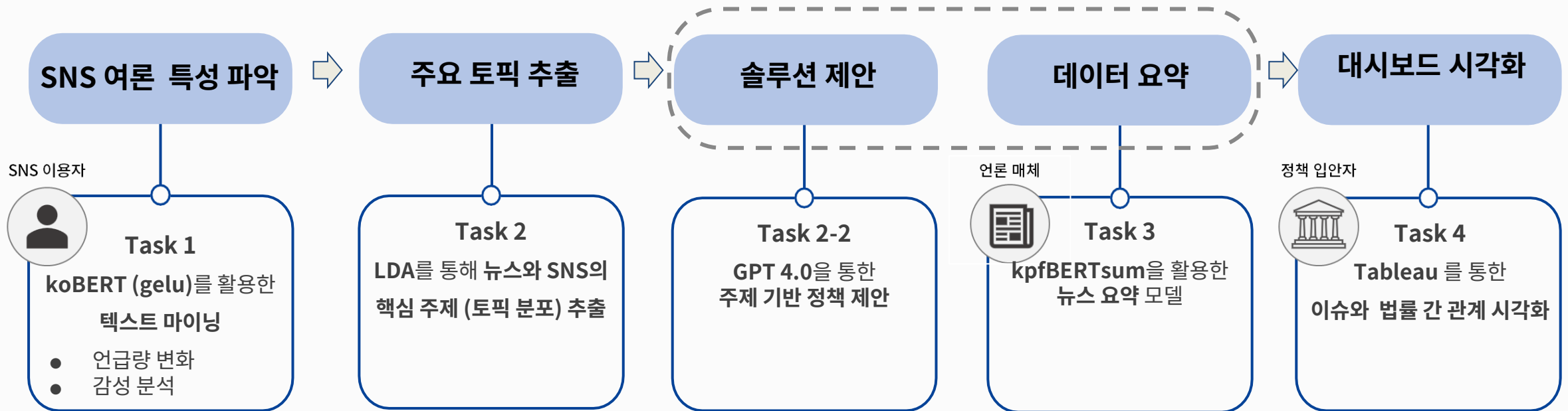
##### 3 AI Report 대시보드 활용 분석 접근성 향상

분석 접근성 제고 및 '여론, 법률, 정책' 관계 이해도 향상 기여

### M'AI Report

= My + AI, 나를 위한 AI 여론 분석 리포트

본 프로젝트는 인공지능 기술을 통해 효과적인 SNS 여론 특성 파악 및 분석 후 ① 여론을 취합한 생성형 AI 기반 법안 제안 ② 뉴스 요약 모델을 통해 『SNS & 언론매체 기반 생성형 AI 리포트』를 제안



# Outline

## 1-3 목표 결과물



국민

“정책 체감도, 실효성 제고 희망”

국민 여론이 때에 맞게, 충분히 반영되었으면 좋겠어요

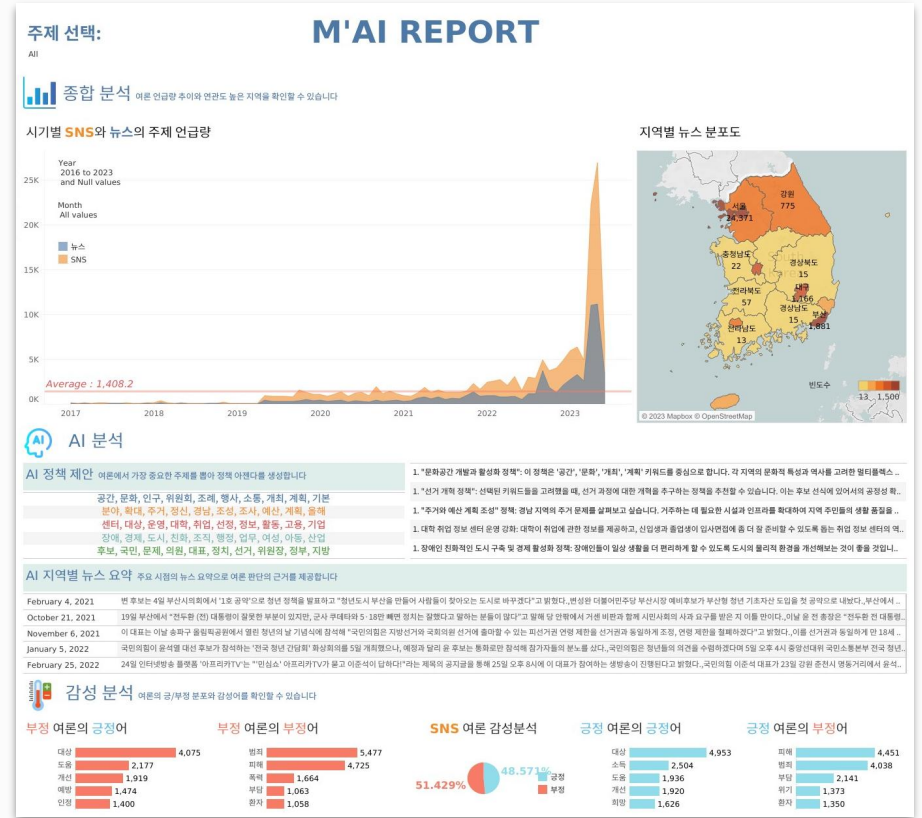
“유익한 여론 정보 파악 희망”

방대한 여론 데이터 속 국민의 의견을 알기 쉽고 신속하게 파악해서 정책 구상에 활용하고 싶어요



국회

여론과 정책의 연결 도구 필요



AI 여론 리포트 기획

02

Data  
활용 데이터

### 데이터 수집 기간 / 채널

**출처:** 2023 DATA·AI 분석 경진대회, 국회도서관 데이터융합분석과

**주제:** 4대 주요 입법 관련 주제  
(스토킹 처벌법, 전세사기, 청년정책, 간호법)

**기간:** 2013~2023(주제 별 상이, 평균 4년 분량)

**채널:** 온라인 뉴스기사, 온라인 커뮤니티, 블로그, Twitter, Instagram, YouTube

### 사용된 데이터

온라인뉴스기사(67,038), 블로그(55,186), **Twitter** (53,192), **온라인 커뮤니티**(31,667), **Instagram**(12,672) \*YouTube 제외  
총 219, 755 건

### 추가 데이터

AI - HUB 한국어 요약 데이터셋: 뉴스요약모델 (kpfBERTsumm) 학습용

국가법령정보센터: 법률 제정/개정 정보용



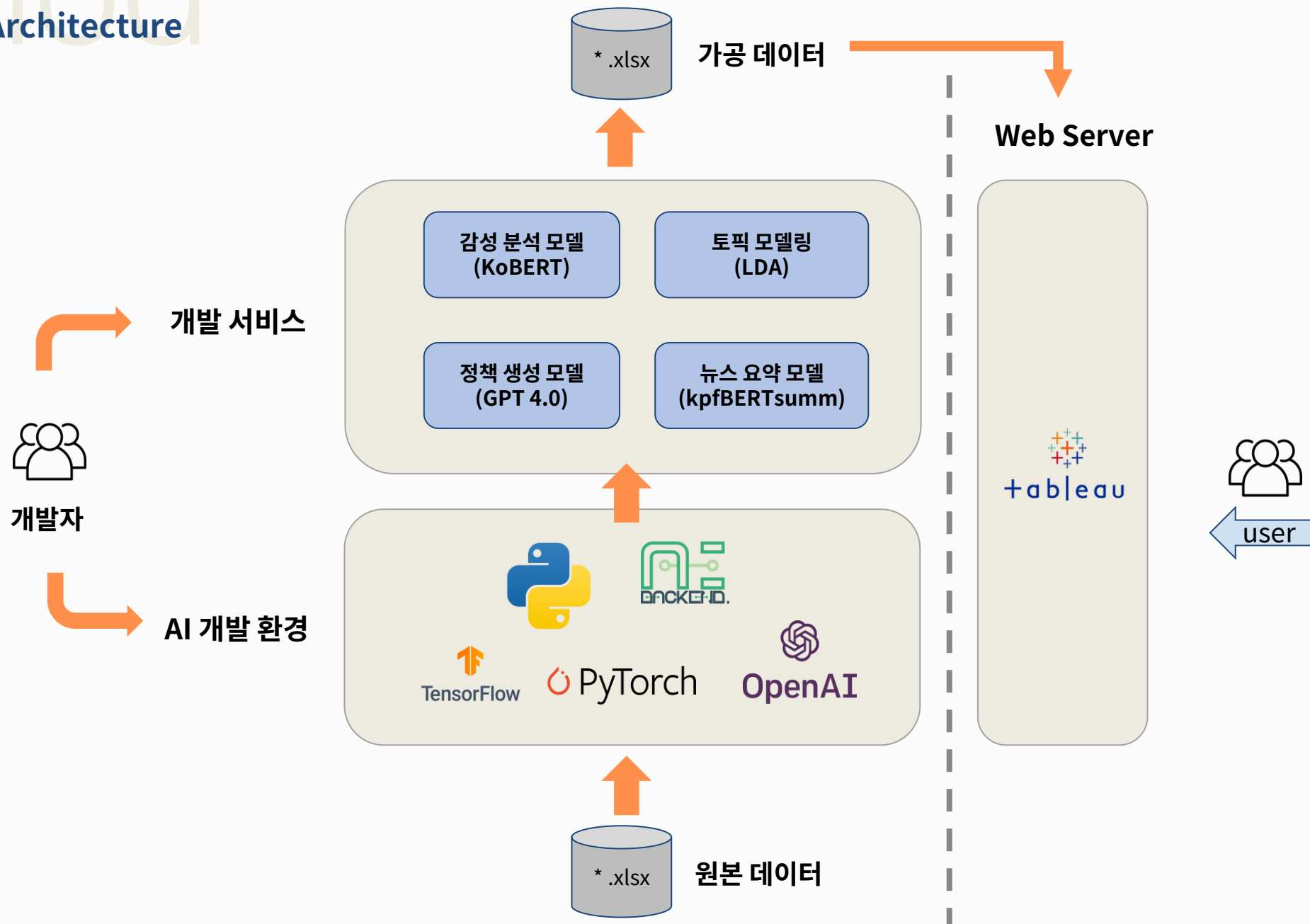
03

# Method

**모델 개발 방법**

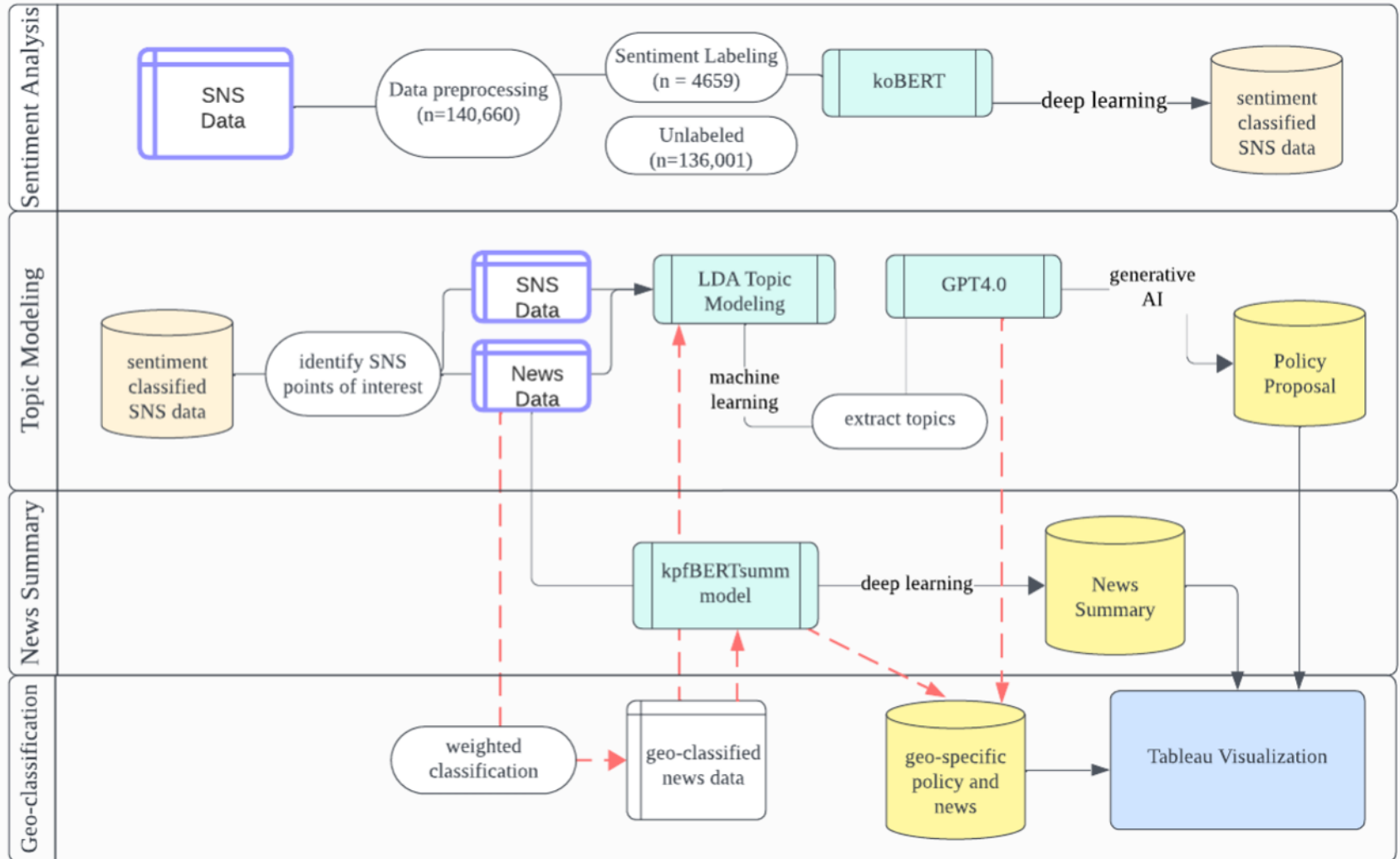
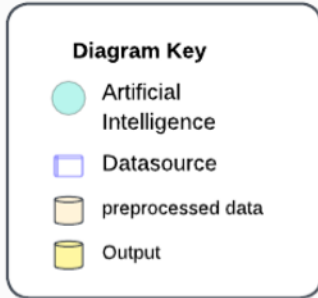
# Method

## 3-1 Project Architecture



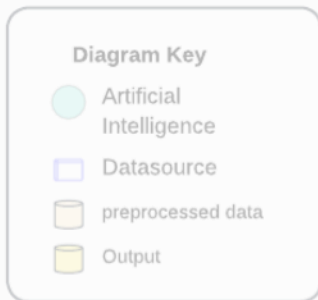
# Method

## 3-2 서비스별 적용 기술

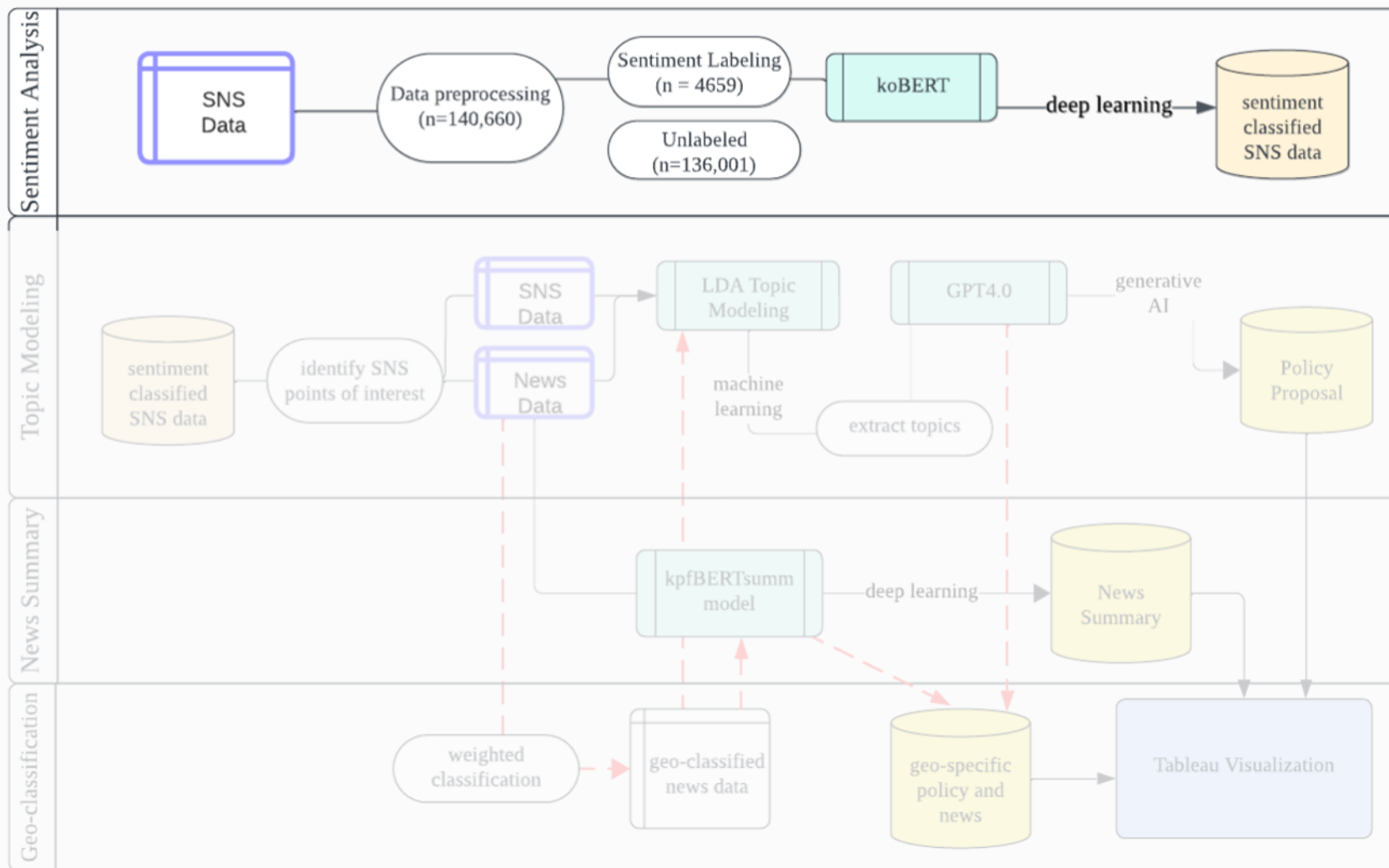


# Method

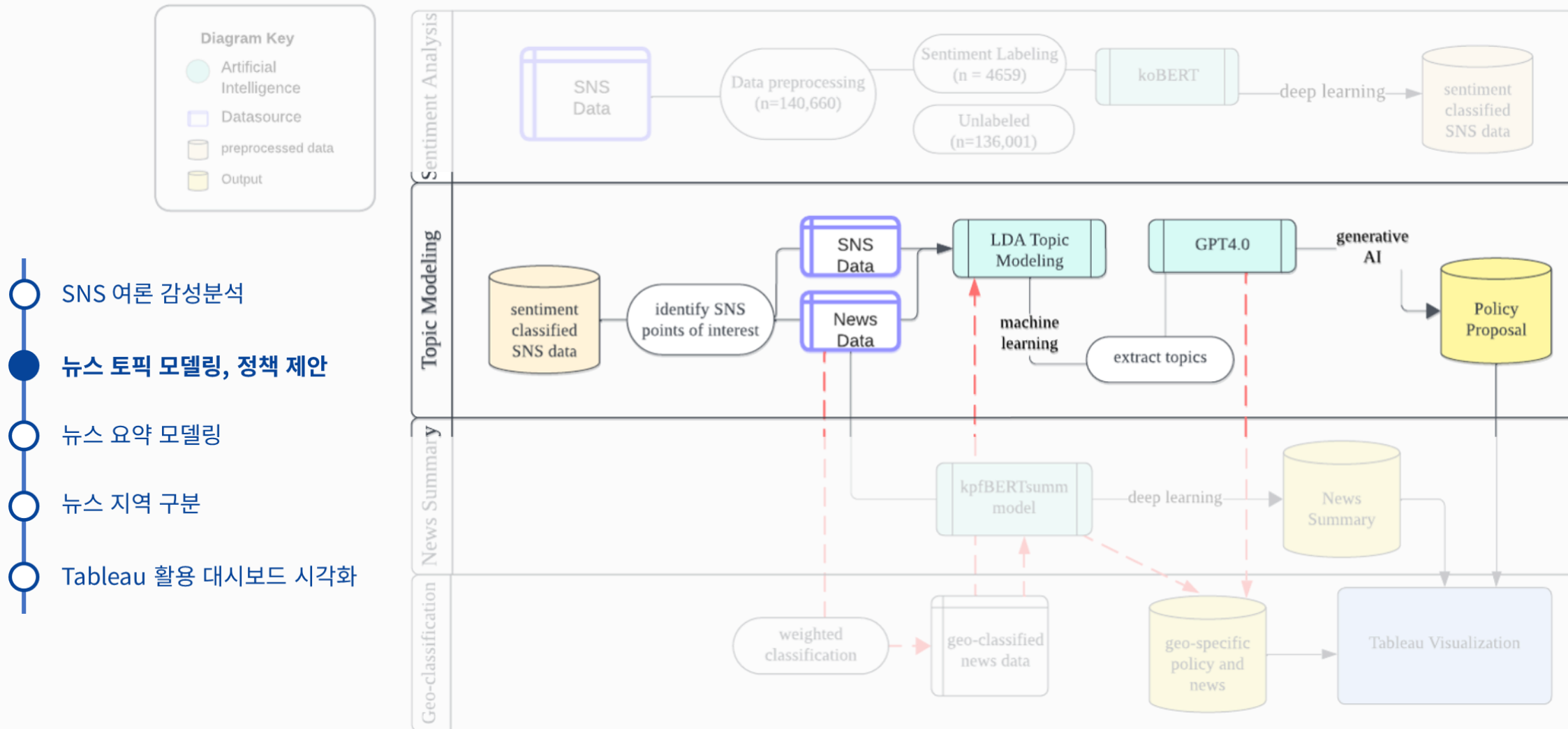
## Task 1: SNS 기반 긍/부정 여론 감성분석 및 분류



- SNS 여론 감성분석
- 뉴스 토픽 모델링, 정책 제안
- 뉴스 요약 모델링
- 뉴스 지역 구분
- Tableau 활용 대시보드 시각화

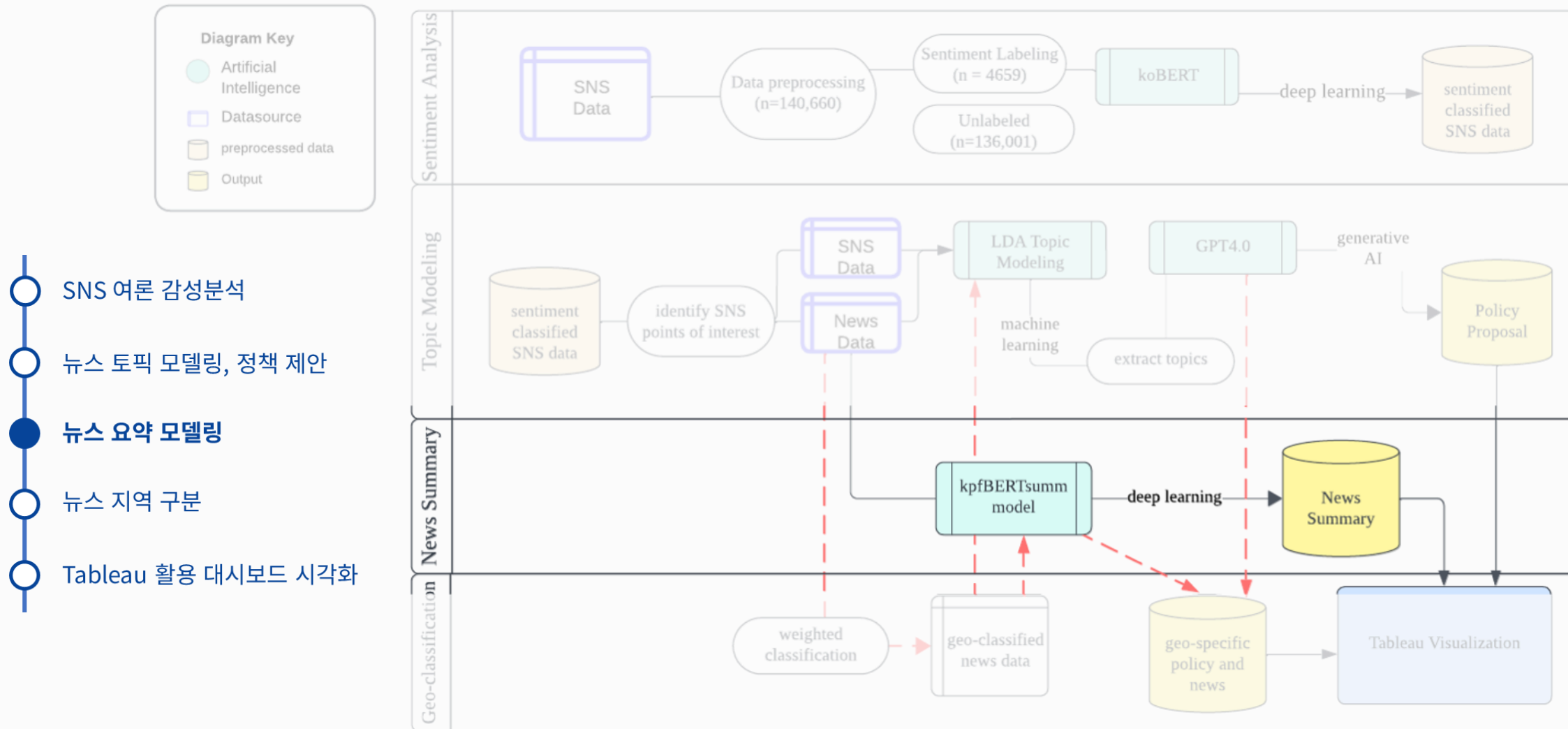


## Task 2: LDA 기반 주요 뉴스 토픽 모델링, GPT 4.0 활용 정책 제안



# Method

## Task 3: kpfBERTsumm 활용 뉴스 요약 모델링



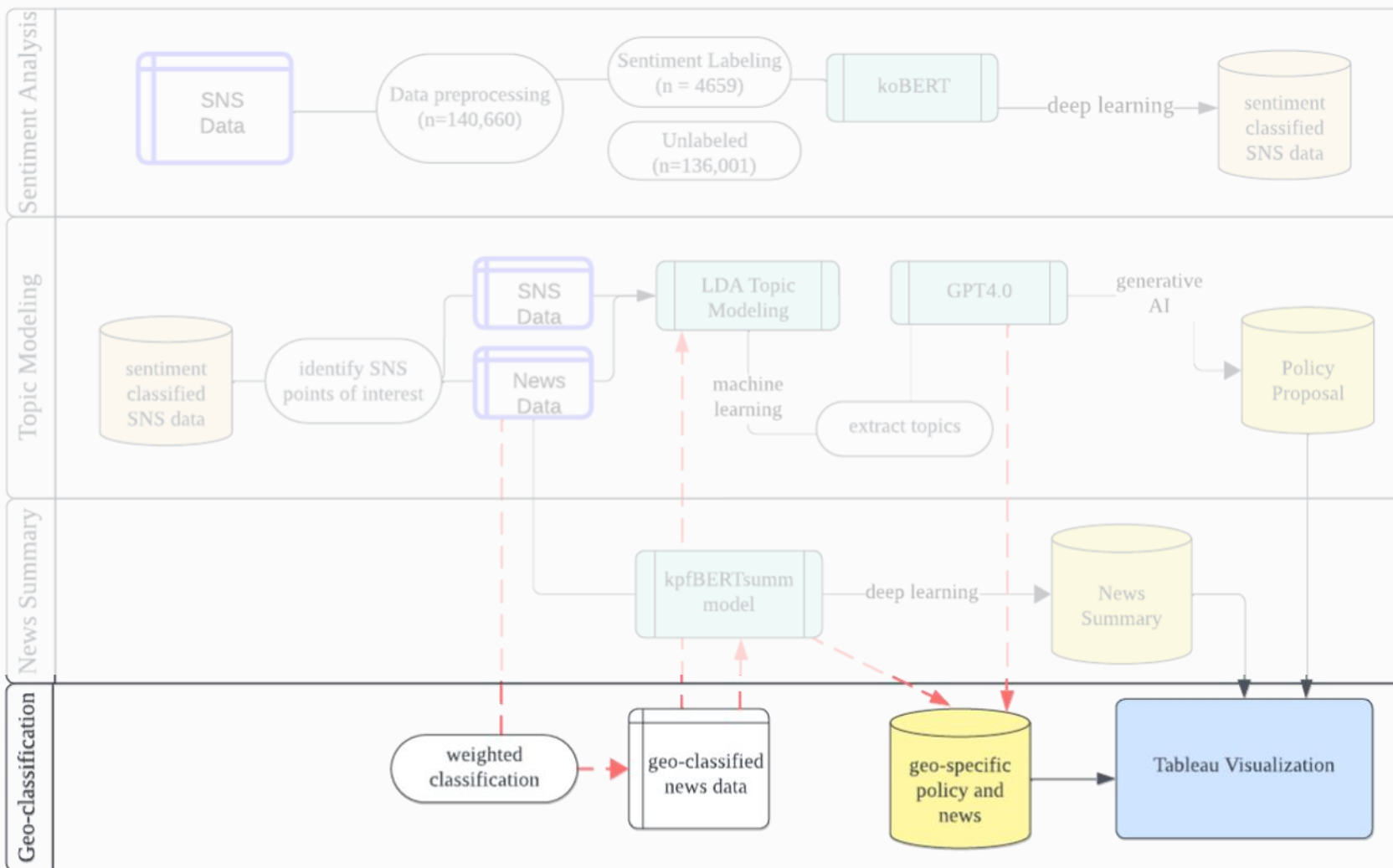
# Method

## Task 4: 지역별 뉴스 분석

**Diagram Key**

- Artificial Intelligence
- Datasource
- preprocessed data
- Output

- SNS 여론 감성분석
- 뉴스 토픽 모델링, 정책 제안
- 뉴스 요약 모델링
- 뉴스 지역 구분
- Tableau 활용 대시보드 시각화



04

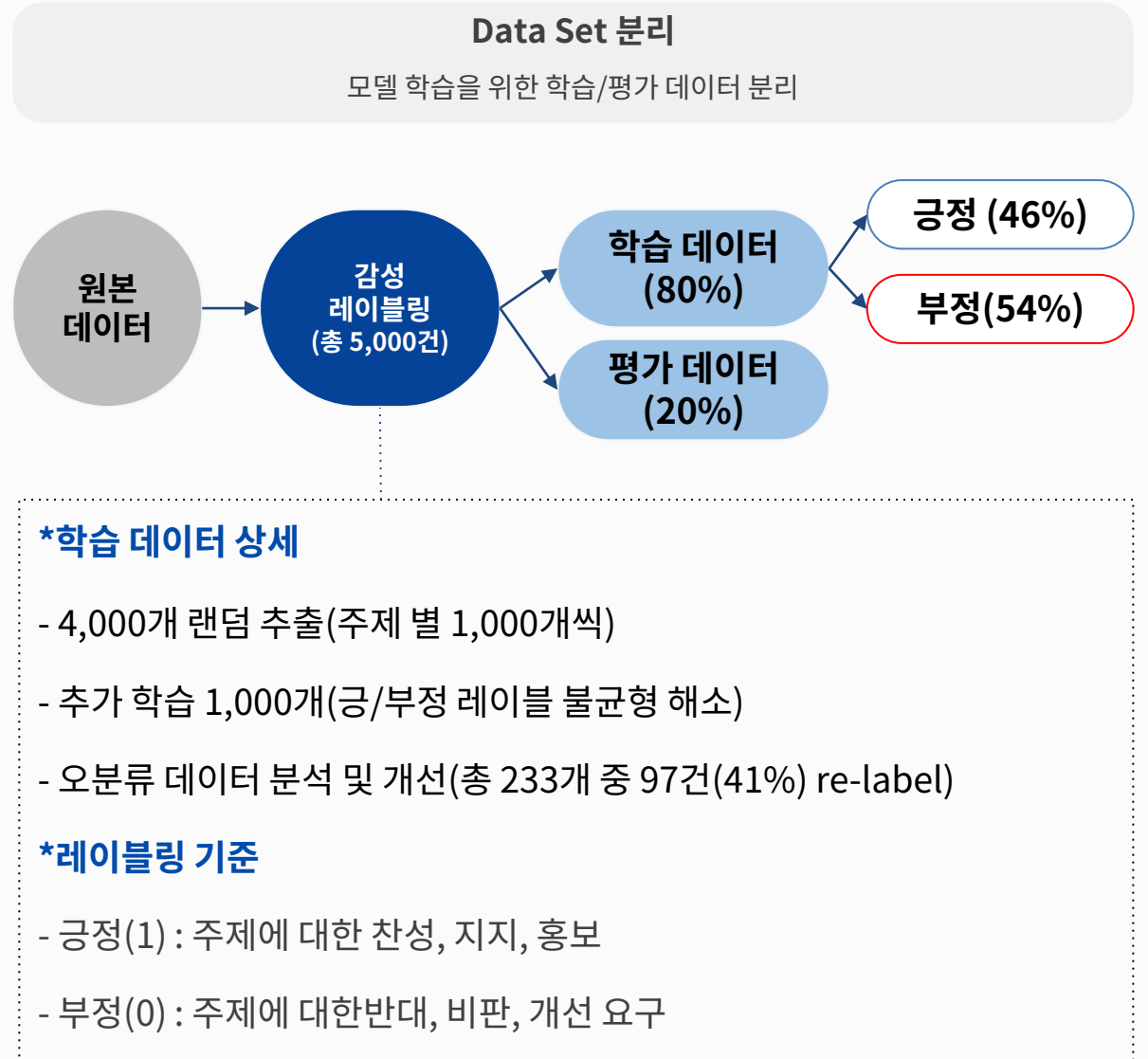
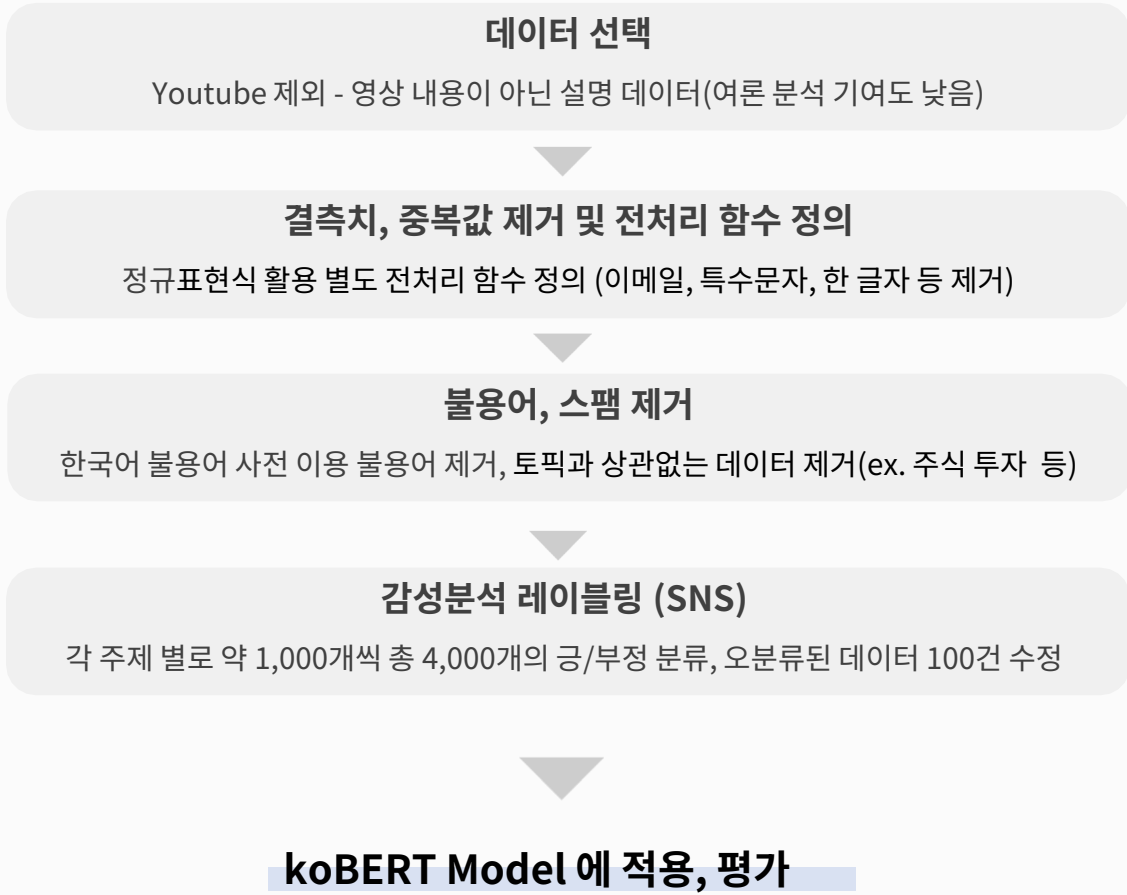
# Analysis

실험 및 평가



# Analysis

## 4-1 데이터 전처리

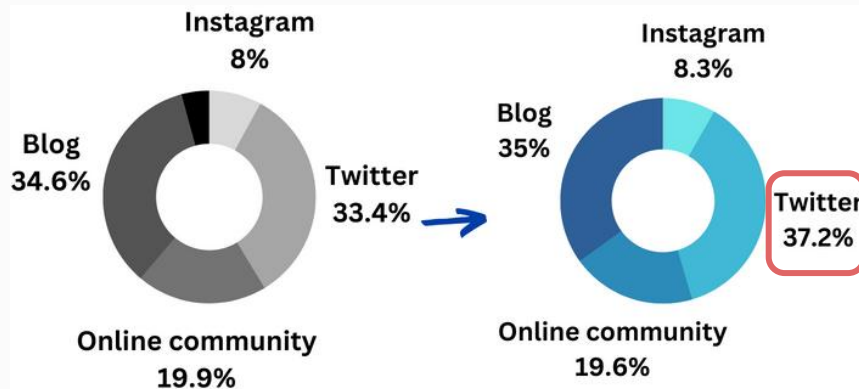
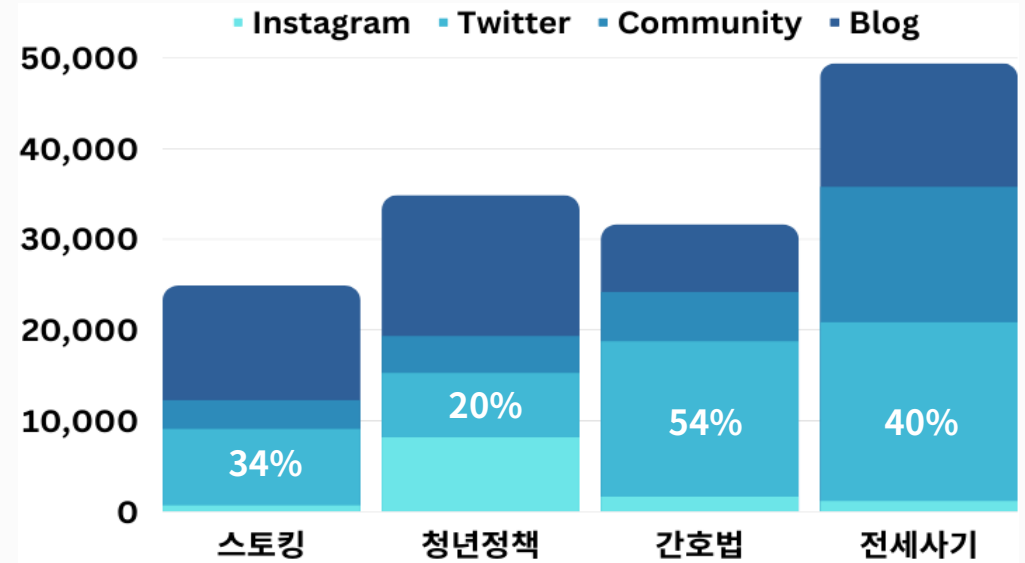


## 4-2 EDA 결과 보고

[ EDA 결론 ] : 전처리 후 데이터 감소율이 가장 낮은 트위터가 ‘유의미한 여론 파악’의 채널이라고 판단  
 → 모델링 활용(추가 학습/보완 시, 트위터 데이터를 중심으로 반영)

### 활용 데이터

	Raw data	Final data(전처리 진행)
Youtube	6,717	N/A
Instagram	12,672	11,626(-8%)
Twitter	53,192	52,286(-2%)
Blog	55,186	49,159(-11%)
Community	31,667	27,589(-13%)
TOTAL	159,434	140,660(-12%)



### Insight

- 전처리 후 트위터 비중 증가(blog, community 데이터 스팸 多)
- 데이터 수집 기간(1년)이 가장 짧은 ‘전세사기’의 데이터 비중 大 (관심 연령층의 범위가 넓다 = 여론 트렌드 파악)
- 토픽 별 여론이 형성되는 채널의 종류와 데이터 비중의 차이 확인

## Task 1: SNS 기반 긍/부정 여론 감성분석 및 분류

### koBERT (gelu) 모델이란?

- SKT Brain KoBERT Model는 BERT의 한국어 버전이며, 트랜스포머를 사용한 한국어의 분석, 이해, 활용에 특화된 양방향성 딥러닝 언어모델
- GeLu(활성화 함수)를 통해 더 높은 정확도를 얻을 수 있다.
- 환경: 4GB NVIDIA GPU, 4GB HIGH RAM

### \*GeLu

Sigmoid와 ReLu의 장점을 결합한 활성화 함수로, 출력 범위는 0~1 사이

### 감성분석 모델 파이프라인

Step 1. 감성 분석을 위한 data preprocessing  
데이터 전처리 및 레이블링

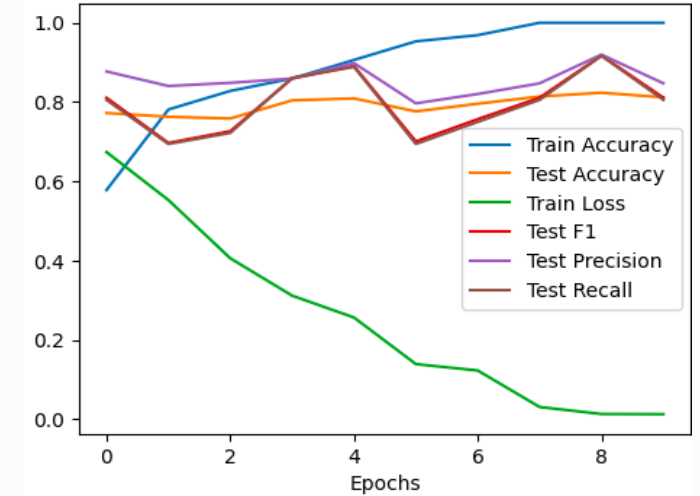
Step 2. koBERT, biLSTM, GRU 모델 학습

Step 3. 모델 간 성능 평가  
정확도, 재현율, F1 - score 활용

Step 4. SNS 긍/부정 여론 구분  
koBERT 딥러닝 모델 활용

토픽 모델링, 뉴스 요약 모델에 활용

### 성능 평가



### [hyper-parameter tuning]

- Epoch : 10번 수행
- batch\_size : 각 데이터의 사이즈 64
- learning rage: 5e-5(약 0.006)

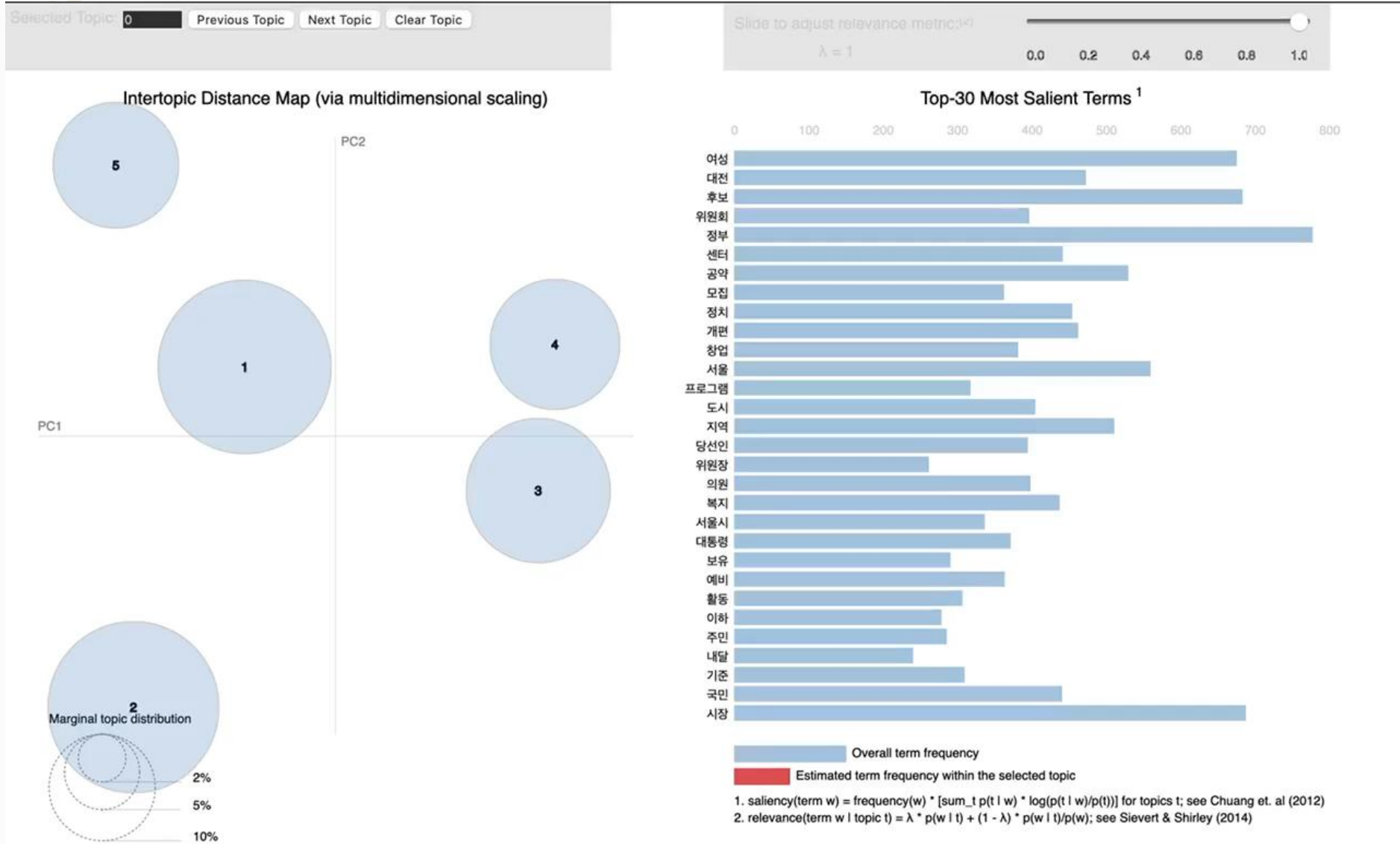
### [Results]

Train 정확도 98.8%      Test 정확도 **82.3%**  
Recall 91.6%              Precision 91.9%  
F1 score **91.7%**

다른 모델에 비해 높은 예측률 → 모델 선정



# Task 2: LDA 토픽 모델링 결과 보고



# Task 3: kpfBERTsumm 활용 뉴스 요약 모델링

## kpfBERTsumm 모델이란?

●한국언론진흥재단에서 구축한 방대한 뉴스기사 코퍼스로 학습한 모델 (kpfBERT)을 이용하여 한국어 뉴스기사 요약에 특화된 모델이다.

●Library: pyToch, Transformer

## 목표

여론이 뒤바뀐 (변곡점) 시기의 관련 뉴스 기사를 추출하고 요약하여 **어떻게 언론매체가 해당 주제를 인식**하는지 신속하고 간략하게 이해할 수 있다.

●변곡점 기반 +/- 1 day의 댓글 수가 많은 기사 3개씩 추출 및 요약

## 뉴스 요약 모델 파이프라인

Step 1. kpfBERTsumm 뉴스 요약 모델 학습

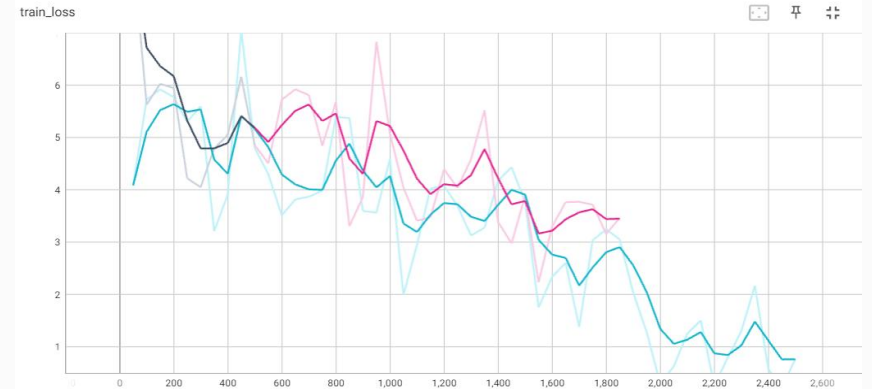
Step 2. 성능 평가  
loss 함수 활용

Step 3. 변곡점 시기의 관련 뉴스 요약

AI Report

## 성능 평가

```
from transformers import BertTokenizerFast, BertModel  
  
BERT_MODEL_NAME = "jinmang2/kpfbert"  
tokenizer = BertTokenizer.from_pretrained(BERT_MODEL_NAME)
```



Test metric	DataLoader 0
test_epoch_average test_loss	11.81667423248291 11.816675186157227

pre-train 학습 데이터: AI-HUB 문서요약 데이터셋  
-원문데이터 30만 건  
-Train : 2000건, Test: 200건, Validation: 200건

Epoch : 5번 수행

batch\_size : 각 데이터의 사이즈 4

**Test loss = 11% >>> 모델 선정**

## Task 3: kpfBERTsumm 결과 보고

### 뉴스 요약 모델 예시

#### #테스트 문장 입력

```
test1 = '''이재명 더불어민주당 대선후보는 26일 변호사비 대납 의혹과 관련, "내가 정말로 변호사비를 불법으로 받았으면 나를 구속하  
이 후보는 이날 오후 전남 신안군 응급의료 전용헬기 계류장에서 열린 '국민반상회' 후 기자들과 만나 한 시민단체 대표가 고액 수임료  
앞서 이민구 깨어있는시민연대당 대표는 이 후보가 특정 변호사에게 수임료로 현금과 주식 등 20억원을 줬다는 의혹을 주장하며 녹취록  
이에 대해 이 후보는 "그것도 조직폭력배 조작에 버금가는 조작사건이라는 게 곧 드러날 것"이라며 "팩트확인을 하고 언급하면 좋겠다.  
그는 "사실이 아니면 무고하고 음해하는 사람들을 무고 혐의나 공직선거법 위반으로 빨리 처리해서 처벌하시라"며 "선거 국면에서 하루  
이어 "허위사실이 드러났으면 당연히 다시는 그런 일이 없게 해야 하는 것 아닌가. 이해가 안 된다"며 "선거관리, 또는 범죄를 단속하는  
이 후보는 또 자신이 구민주-동교동계와 접촉해 복당을 타진했다는 언론보도와 관련해선 "구체적으로 어떤 사람을 범주별로 나눠 무슨  
종전에 언급했던 '대사면' 방침을 재확인한 셈이다. 그는 "민주당에 계셨던 분, 또 민주당에 있지 않았더라도 앞으로 함께할 분들에게  
그는 "현재 민주당이 이미 열린민주당과의 통합을 협의하고 있다"며 "거기에 더해서 꼭 민주계라고 말할 필요는 없고 부패사범이나 파  
언론보도에 따르면, 이 후보는 최근 구민주계인 정대철 전 고문과 연락을 주고 받으며 천정배, 정동영 전 의원 등 민주당을 탈당했던 9  
...  
'''
```

```
rtn = summarize_test(test1)
```

\*\*\* 입력한 문단의 요약문은 ...

- [ 1 ] 이재명 더불어민주당 대선후보는 26일 변호사비 대납 의혹과 관련, "내가 정말로 변호사비를 불법으로 받았으면 나를 구속하라"고
- [ 2 ] 이 후보는 이날 오후 전남 신안군 응급의료 전용헬기 계류장에서 열린 '국민반상회' 후 기자들과 만나 한 시민단체 대표가 고액 수임료
- [ 3 ] 이에 대해 송평수 선대위 부대변인은 "허위사실"이라며 "깨시민당 이 대표에게 제보를 했다는 시민단체 대표 이모 씨가 제3자로부

## Task 4: 지역별 뉴스 분석

- SNS 여론 감성분석
- 뉴스 토픽 모델링, 정책 제안
- 뉴스 요약 모델링
- 뉴스 지역 구분**
- Tableau 활용 대시보드 시각화

### 파이프라인

#### Step 1. 지역 분류(가중치 적용)\*

\*지역 구분 기준 : '작성자', '제목', '내용'열에 언급된 지역명으로 구분

- 작성자가 지역신문사인 경우 : 해당 지역으로 구분
- 작성자가 지역신문사가 아닌 경우 : 제목, 내용에 언급된 지역으로 구분 \*단, 제목은 가중치 2배로 계산

#### Step 2. 지역 뉴스 요약

수도권 제외 뉴스 요약

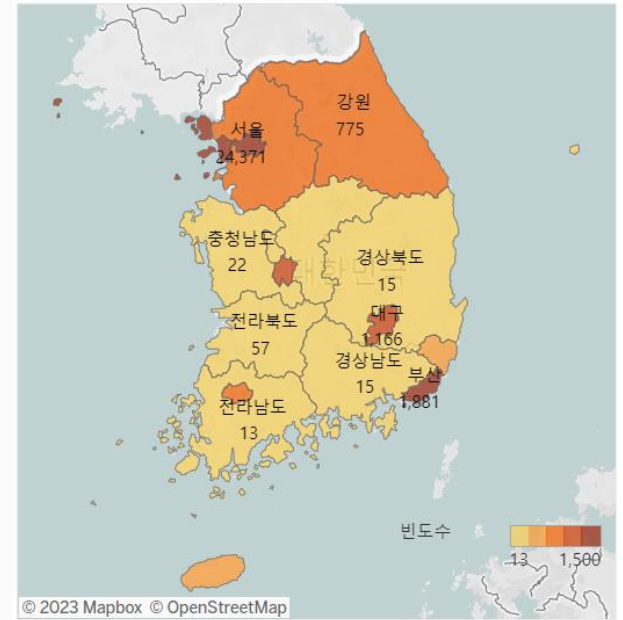
#### Step 3. 지역 뉴스 토픽 모델링

#### Step 4. 생성형 AI 기반 지역 정책 추천

뉴스 토픽 모델링 기반 지역 맞춤형 정책 방향성 제공

### 적용 예시

지역별 뉴스 분포도





Visualize

05

**활용 계획 및 기대효과**

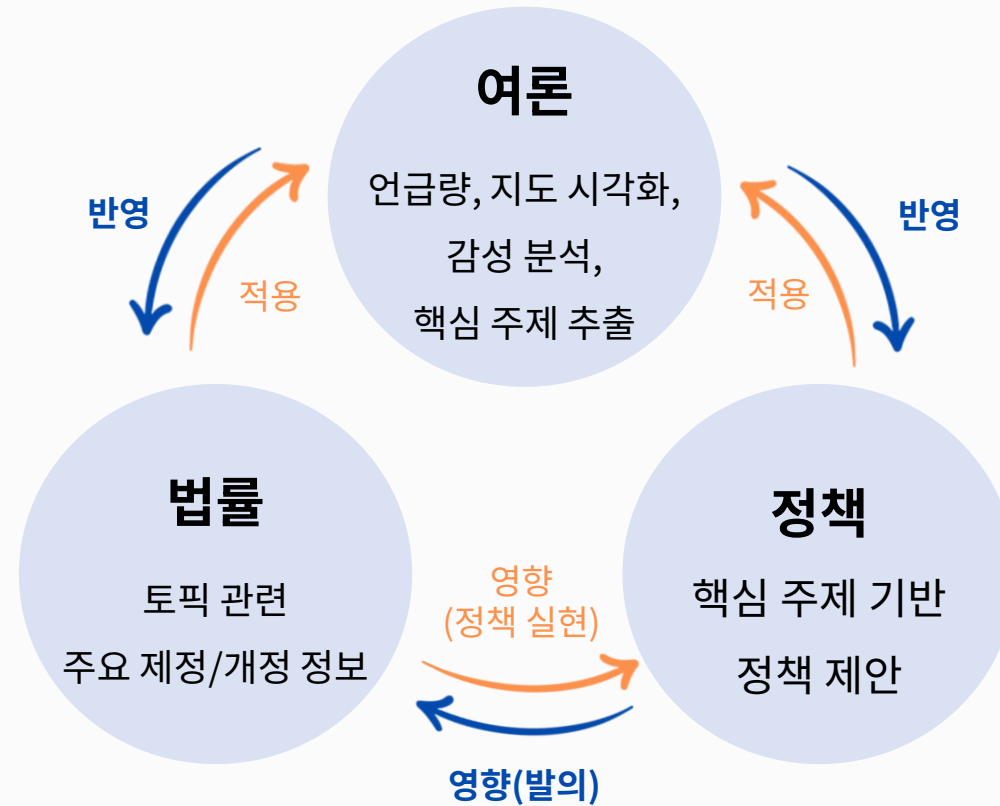
# M'AI Report

My + AI

**= 나를 위한 AI 여론 분석 리포트**

Tableau(데이터 시각화 소프트웨어) 대시보드 구현을 통한

# 여론, 정책, 법률 이해도 제고



- SNS 여론 감성분석
- 뉴스 토픽 모델링, 정책 제안
- 뉴스 요약 모델링
- 뉴스 지역 구분
- Tableau 활용 대시보드 시각화

## 5-3 결과물 소개

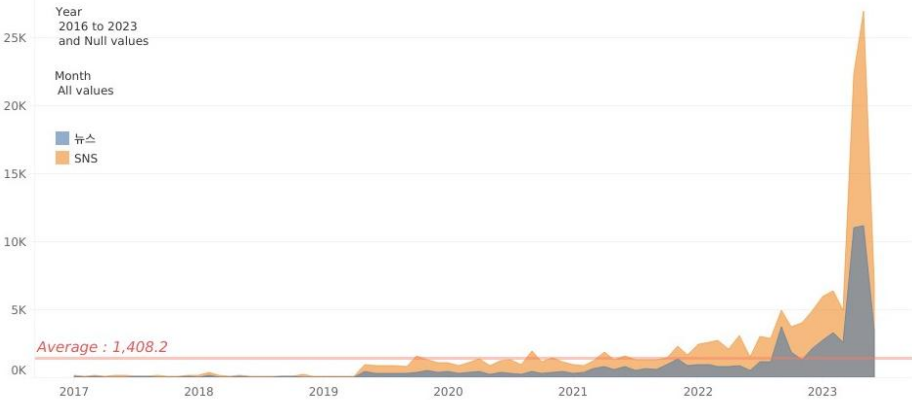
주제 선택:

### M'AI REPORT

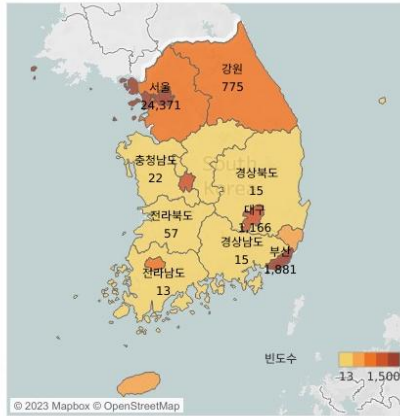
All

**종합 분석** 여론 언급량 추이와 연관도 높은 지역을 확인할 수 있습니다

시기별 SNS와 뉴스의 주제 언급량



지역별 뉴스 분포도



### AI 분석

**AI 정책 제안** 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

공간, 문화, 인구, 위원회, 조례, 행사, 소통, 개최, 계획, 기본 분야, 확대, 주거, 정신, 경남, 조성, 조사, 예산, 계획, 올해 센터, 대상, 운영, 대학, 취업, 선정, 정보, 활동, 고용, 기업 장애, 경제, 도시, 친화, 조직, 행정, 업무, 여성, 아동, 산업 후보, 국민, 문제, 의원, 대표, 정치, 선거, 위원장, 정부, 지방

- 1. "문화공간 개발과 활성화 정책": 이 정책은 '공간', '문화', '개회', '계획' 키워드를 중심으로 합니다. 각 지역의 문화적 특성과 역사를 고려한 멀티플렉스...
- 1. "선거 개혁 정책": 선출된 키워드들을 고려했을 때, 선거 과정에 대한 개혁을 추구하는 정책을 추천할 수 있습니다. 이는 후보 선시에 있어서의 공정성 확...
- 1. "주거와 예산 계획 조성" 정책: 경남 지역의 주거 문제를 살펴보고 싶습니다. 거주하는 데 필요한 시설과 인프라를 확대하여 지역 주민들의 생활 품질을...
- 1. 대학 취업 정보 센터 운영 강화: 대학이 취업에 관한 정보를 제공하고, 산업생과 졸업생이 입사면접에 좀 더 잘 준비할 수 있도록 돕는 취업 정보 센터의 역...
- 1. 장애인 친화적인 도시 구축 및 경제 활성화 정책: 장애인들이 일상 생활을 더 편리하게 할 수 있도록 도시의 물리적 환경을 개선해보는 것이 좋을 것입니...

**AI 지역별 뉴스 요약** 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

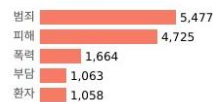
- February 4, 2021 변 후보는 4일 부산시의회에서 '1호 공약'으로 청년 정책을 발표하고 "청년도시 부산을 만들어 사람들이 찾아오는 도시로 바꾸겠다"고 밝혔다. 변성원 더불어민주당 부산시장 예비후보가 부산형 청년 기초자선 도입을 첫 공약으로 내놨다. 부산에서..
- October 21, 2021 19일 부산에서 "전두환 (전) 대통령이 잘못된 부분이 있지만, 군사 쿠데타와 5-18만 배면 정치는 잘못다고 말하는 분들이 많다"고 말해 당 안팎에서 거센 비판과 함께 시민사회의 사과 요구를 받은 지 이틀 만이다. 이날 윤 전 총장은 "전두환 전 대통령..
- November 6, 2021 이 대표는 이날 송파구 올림픽공원에서 열린 청년의 날 기념식에 참석해 "국민의힘은 지방선거와 국회의원 선거에 출마할 수 있는 피선거권 연령 제한을 선거권과 동일하게 조정, 연령 제한을 철폐하겠다"고 밝혔다. 이를 선거권과 동일하게 만 18세..
- January 5, 2022 국민의힘이 윤석열 대선 후보가 참석하는 '전국 청년 간담회' 화상회의를 5일 개최했으나, 예정과 달리 윤 후보는 통화로만 참석해 참가자들의 분노를 샀다. 국민의힘은 청년들의 의견을 수렴하겠다고며 5일 오후 4시 중앙선대위 국민소통본부 전국 청년..
- February 25, 2022 24일 인터넷방송 플랫폼 '아프리카TV'는 "'민심쇼' 아프리카TV가 묻고 이준석이 답하다"라는 제목의 공지글을 통해 25일 오후 8시에 이 대표가 참여하는 생방송이 진행된다고 밝혔다. 국민의힘 이준석 대표가 23일 강원 춘천시 명동거리에서 윤석...

**감성 분석** 여론의 긍/부정 분포와 감성어를 확인할 수 있습니다

부정 여론의 긍정어



부정 여론의 부정어



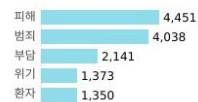
SNS 여론 감성분석



긍정 여론의 긍정어



긍정 여론의 부정어



## [ 종합 페이지 ]

### 1. 종합 분석

- **뉴스/SNS 언급량 추이**  
시계열 그래프를 통해 뉴스/SNS 언급량 수치 변화
- **지역 연관 뉴스 정보 및 정책 제안(지역 특성화)**  
뉴스 지역 구분\* 및 지도 시각화

\*지역 구분 기준 : '작성자', '제목', '내용' 열에 언급된 지역명으로 구분  
- 작성자가 **지역신문사인 경우** : 해당 지역으로 구분  
- 작성자가 **지역신문사가 아닌 경우** : 제목, 내용에 언급된 지역으로 구분 (단, 제목은 가중치 2배로 계산)

## 5-3 결과물 소개

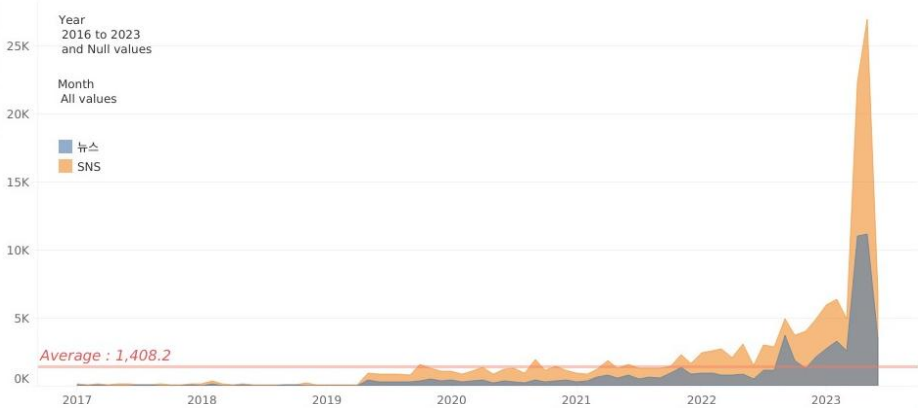
주제 선택:

### M'AI REPORT

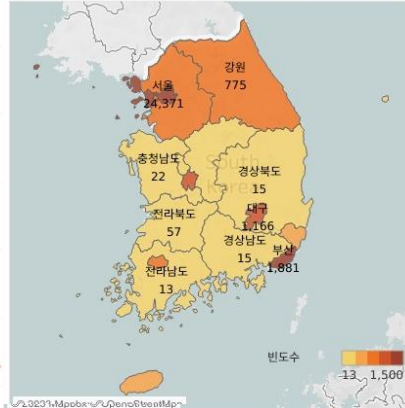
All

**종합 분석** 여론 언급량 추이와 연관도 높은 지역을 확인할 수 있습니다

시기별 SNS와 뉴스의 주제 언급량



지역별 뉴스 분포도



### AI 분석

**AI 정책 제안** 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

공간, 문화, 인구, 위원회, 조례, 행사, 소통, 개최, 계획, 기본 분야, 확대, 주거, 장신, 경남, 조성, 조사, 예산, 계획, 올해 센터, 대상, 운영, 대학, 취업, 선정, 정보, 활동, 고용, 기업 장래, 경제, 도시, 친화, 조직, 행정, 업무, 여성, 아동, 산업 후보, 국민, 문제, 의원, 대표, 정치, 선거, 위원장, 정부, 지방

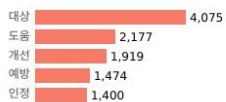
- "문화공간 개발과 활성화 정책": 이 정책은 '공간', '문화', '개척', '계획' 키워드를 중심으로 합니다. 각 지역의 문화적 특성과 역사를 고려한 멀티플렉스...
- "선거 개혁 정책": 선출된 키워드들을 고려했을 때, 선거 과정에 대한 개혁을 추구하는 정책을 추천할 수 있습니다. 이는 후보 선시에 있어서의 공정성 확...
- "주거와 예산 계획 조성" 정책: 경남 지역의 주거 문제를 살펴보고 싶습니다. 거주하는 데 필요한 시설과 인프라를 확대하여 지역 주민들의 생활 품질을 ...
- 대학 취업 정보 센터 운영 강화: 대학이 취업에 관한 정보를 제공하고, 신입생과 졸업생이 입사면접에 좀 더 잘 준비할 수 있도록 돕는 취업 정보 센터의 역...
- 장애인 친화적인 도시 구축 및 경제 활성화 정책: 장애인들이 일상 생활을 더 편리하게 할 수 있도록 도시의 물리적 환경을 개선해보는 것이 좋을 것입니...

**AI 지역별 뉴스 요약** 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

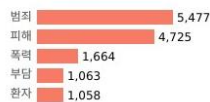
February 4, 2021	변 후보는 4일 부산시의회에서 '1호 공약'으로 청년 정책을 발표하고 "청년도시 부산을 만들어 사람들이 찾아오는 도시로 바꾸겠다"고 밝혔다. 변성완 더불어민주당 부산시장 예비후보가 부산형 청년 기초자산 도입을 첫 공약으로 내놨다. 부산에서 ...
October 21, 2021	19일 부산에서 "전두환 (전) 대통령이 잘못된 부분이 있지만, 군사 쿠데타와 5·18만 빼면 정치는 잘못했다고 말하는 분들이 많다"고 말해 당 안팎에서 거센 비판과 함께 시민사회의 사과 요구를 받은 지 이틀 만이다. 이날 윤 전 총장은 "전두환 전 대통령 ...
November 6, 2021	이 대표는 이날 송파구 올림픽공원에서 열린 청년의 날 기념식에 참석해 "국민의힘은 지방선거와 국회의원 선거에 출마할 수 있는 피선거권 연령 제한을 선거권과 동일하게 조정, 연령 제한을 철폐하겠다"고 밝혔다. 이를 선거권과 동일하게 만 18세 ...
January 5, 2022	국민의힘이 윤석열 대선 후보가 참석하는 '전국 청년 간담회' 화상회의를 5일 개최했다. 예정과 달리 윤 후보는 통화로만 참석해 참가자들의 분노를 샀다. 국민의힘은 청년들의 의견을 수렴하겠다고 5일 오후 4시 중앙선대위 국민소통본부 전국 청년 ...
February 25, 2022	24일 인터넷방송 플랫폼 '아프리카TV'는 "'민심소' 아프리카TV가 묻고 이준석이 답하다"라는 제목의 공지글을 통해 25일 오후 8시에 이 대표가 참여하는 생방송이 진행된다고 밝혔다. 국민의힘 이준석 대표가 23일 강원 춘천시 명동거리에서 윤석...

**감성 분석** 여론의 긍/부정 분포와 감성어를 확인할 수 있습니다

부정 여론의 긍정어



부정 여론의 부정어



SNS 여론 감성분석



긍정 여론의 긍정어



긍정 여론의 부정어



## [ 종합 페이지 ]

### 1. 종합 분석

- 뉴스/SNS 언급량 추이  
시계열 그래프를 통해 뉴스/SNS 언급량 수치 변화
- 지역 연관 뉴스 정보 및 정책 제안(지역 특성화)  
뉴스 지역 구분\* 및 지도 시각화

\*지역 구분 기준 : '작성자', '제목', '내용' 열에 언급된 지역명으로 구분  
- 작자가 지역신문사인 경우 : 해당 지역으로 구분  
- 작자가 지역신문사가 아닌 경우 : 제목, 내용에 언급된 지역으로 구분 (단, 제목은 가중치 2배로 계산)

### 2. AI 분석

- AI 정책 제안  
여론에서 가장 중요한 주제를 뽑아 정책 아젠다 생성
- AI 지역별 뉴스 요약  
주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공

## 5-3 결과물 소개

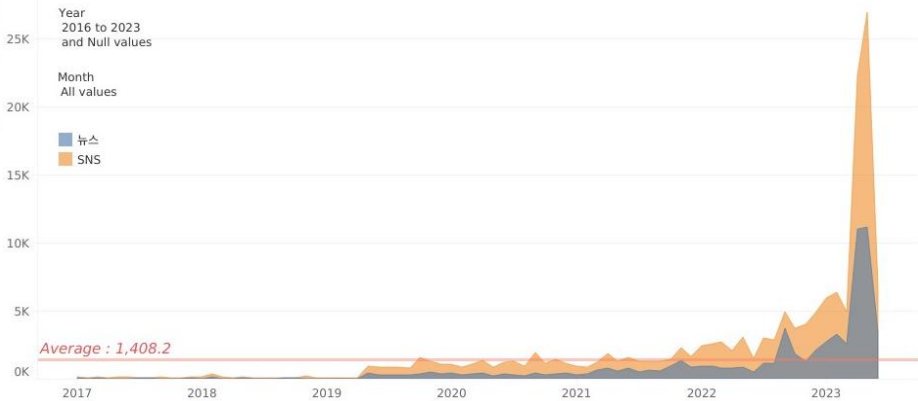
주제 선택:

### M'AI REPORT

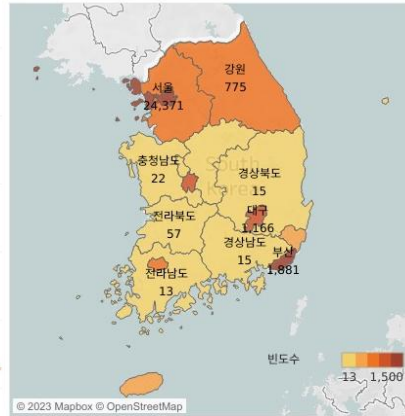
All

**종합 분석** 여론 언급량 추이와 연관도 높은 지역을 확인할 수 있습니다

시기별 SNS와 뉴스의 주제 언급량



지역별 뉴스 분포도



**AI 분석**

**AI 정책 제안** 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

공간, 문화, 인구, 위원회, 조례, 행사, 소통, 개최, 계획, 기본 분야, 확대, 주거, 정신, 경남, 조성, 조사, 예산, 계획, 올해 센터, 대상, 운영, 대학, 취업, 선정, 정보, 활동, 고용, 기업 장애, 경제, 도시, 친화, 조직, 행정, 업무, 여성, 아동, 산업 후보, 국민, 문제, 의원, 대표, 정치, 선거, 위원장, 정부, 지방

1. "문화공간 개발과 활성화 정책": 이 정책은 '공간', '문화', '개회', '계획' 키워드를 중심으로 합니다. 각 지역의 문화적 특성과 역사를 고려한 멀티플렉스...

1. "선거 개혁 정책": 선출된 키워드들을 고려했을 때, 선거 과정에 대한 개혁을 추구하는 정책을 추천할 수 있습니다. 이는 후보 선시에 있어서의 공정성 확...

1. "주거와 예산 계획 조성" 정책: 경남 지역의 주거 문제를 살펴보고 싶습니다. 거주하는 데 필요한 시설과 인프라를 확대하여 지역 주민들의 생활 품질을...

1. 대학 취업 정보 센터 운영 강화: 대학이 취업에 관한 정보를 제공하고, 신입생과 졸업생이 입사면접에 좀 더 잘 준비할 수 있도록 돕는 취업 정보 센터의 역...

1. 장애인 친화적인 도시 구축 및 경제 활성화 정책: 장애인들이 일상 생활을 더 편리하게 할 수 있도록 도시의 물리적 환경을 개선해 보는 것이 좋을 것입니...

**AI 지역별 뉴스 요약** 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

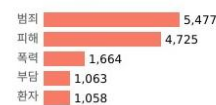
February 4, 2021	변 후보는 4일 부산시의회에서 '1호 공약'으로 청년 정책을 발표하고 "청년도시 부산을 만들어 사람들이 찾아오는 도시로 바꾸겠다"고 밝혔다. "변성원 더불어민주당 부산시장 예비후보가 부산형 청년 기초자산 도입을 첫 공약으로 내놨다. 부산에서..
October 21, 2021	19일 부산에서 "전두환 (전) 대통령이 잘못된 부분이 있지만, 군사 쿠데타와 5-18만 배면 정치는 잘못했다고 말하는 분들이 많다"고 말해 당 안팎에서 거센 비판과 함께 시민사회의 사과 요구를 받은 지 이틀 만이다. 이날 윤 전 총장은 "전두환 전 대통령..
November 6, 2021	이 대표는 이날 송파구 올림픽공원에서 열린 청년의 날 기념식에 참석해 "국민의힘은 지방선거와 국회의원 선거에 출마할 수 있는 피선거권 연령 제한을 선거권과 동일하게 조정, 연령 제한을 철폐하겠다"고 밝혔다. 이를 선거권과 동일하게 만 18세..
January 5, 2022	국민의힘이 윤석열 대선 후보가 참석하는 '전국 청년 간담회' 화상회의를 5일 개최했으나, 예정과 달리 윤 후보는 통화로만 참석해 참가자들의 분노를 샀다. 국민의힘은 청년들의 의견을 수렴하겠다고며 5일 오후 4시 중앙선대위 국민소통본부 전국 청년..
February 25, 2022	24일 인터넷방송 플랫폼 '아프리카TV'는 "민심쇼 아프리카TV가 묻고 이준석이 답하다"라는 제목의 굿즈를 통해 25일 오후 8시에 이 대표가 참여하는 셋박쇼가 진행된다고 밝혔다. 국민의힘 이준석 대표가 23일 강원 춘천시 명동거리에서 유선..

**감성 분석** 여론의 긍/부정 분포와 감성어를 확인할 수 있습니다

부정 여론의 긍정어



부정 여론의 부정어



SNS 여론 감성분석



긍정 여론의 긍정어



긍정 여론의 부정어



## [ 종합 페이지 ]

### 1. 종합 분석

- 뉴스/SNS 언급량 추이  
시계열 그래프를 통해 뉴스/SNS 언급량 수치 변화
- 지역 연관 뉴스 정보 및 정책 제안(지역 특성화)  
뉴스 지역 구분\* 및 지도 시각화

\*지역 구분 기준 : '작성자', '제목', '내용'열에 언급된 지역명으로 구분  
- 작성자가 지역신문사인 경우 : 해당 지역으로 구분  
- 작성자가 지역신문사가 아닌 경우 : 제목, 내용에 언급된 지역으로 구분 (단, 제목은 가중치 2배로 계산)

### 2. AI 분석

- AI 정책 제안  
여론에서 가장 중요한 주제를 뽑아 정책 아젠다 생성
- AI 지역별 뉴스 요약  
주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공

### 3. 감성분석

- 토픽 별 SNS 여론의 긍/부정 분포 및 감성어 상세



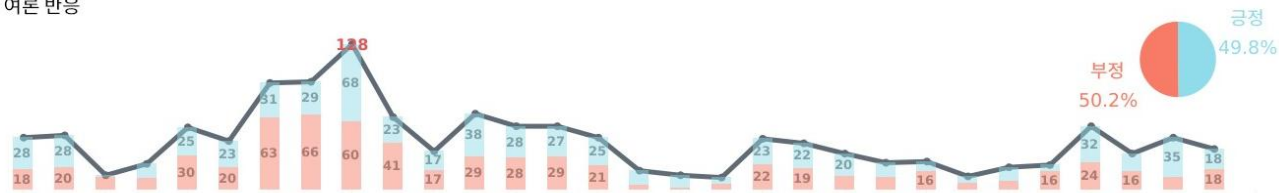
## 5-3 결과물 소개

**주요 시점별 여론 분석** 여론 이해에 필요한 주요 시점을 중심으로 시기별 여론 추이를 확인할 수 있습니다(Point 지점을 클릭 해보세요!)

청년 정책의 시기별 여론 반응



SNS 여론 반응



### AI 분석

AI 정책 제안 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

이달의 정책 주제

제안 내용

청년 구직자 지원금 제도 강화

청년들의 어려운 취업 환경을 겪는 것을 돕기 위해 구직 활동을 지원하는 지원금 제도를 강화할 수 있습니다. 지원금의 지급 규모를 확대하거나 지원 대상을 확대하는 방..

청년 노동자를 위한 개혁 정책

청년 노동자들의 근무환경 개선을 위한 규정 개혁 등 청년 노동자의 권익을 보호하고, 임금 및 기타 근로조건을 개선하는 정책을 추진할 수 있습니다.

청년 소통 센터 구축

청년들이 정치·경제·사회적 이슈에 대해 의견을 나누고, 정보를 얻을 수 있는 플랫폼을 만드는 것이 필요합니다. 이 센터에서는 청년들의 의견이 정책 결정 과정에 반영..

청년 지원 위원회의 설립

청년 문제를 전담하여 다루는 위원회를 설립하는 것이 필요합니다. 이 곳은 청년들이 직면한 문제를 집중적으로 연구하고 해결방안을 마련하는 역할을 하게 됩니다.

AI 뉴스 요약 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

댓글 수(건)

날짜	뉴스1	기사 제목	요약	댓글 수(건)
April 9, 2021	YTN	민주당 2030 의원들의 반성문... '조국 사태' 첫 공개...	"이번 선거에서도 성난 부동산 민심에 공경 가치 문제로 상징되는 LH 사태가 기름을 부었고, 문재인 정부가 표방한 기회의 평등, 과정의 공정, 결과의 정의는 다시 한번 흔들렸습니다. 조국 사태로 공경 가치 문제 본격적으로 불거져 인국공 사태·윤미향·박원순까지 공경 문제 지속 LH 사태로 文 정부 표방 공경 가치 또 흔들려 민주당 의원들, 조국 지키기 첫 공개 비판 영커 더불어민주당 2..	684
	뉴스1	민주당 2030 의원들 "조국 관련, "가르치려 한 오만함, 청년과 단절 원인" 반성 "박원순 피해자 제대로 된 사과와 반성 이뤄지지 않았어" 이소영 의원들 비롯한 더불어민주당 2030의원들이 9일 오전 서울 여의도 국회 소통관에서 사과할 용의 있어... '착각·오만... 더불어민주당 2030의원 입장문 발표를 기자회견을 하고 있다. 민주당 오영환·이소영·장경태·장철민·전용기 의원은 이날 오전 국회 소통관에서 기자회견을 열고 "선거 유세 현장과 삶의 현장..	1,868	
	한겨레	민주당 2030 의원들 그동안 민주당에서 초선 의원들이 당내 현안에 비판적인 입장을 밝힌 것은 이번이 처음으로, 패배 뒤 뒤늦은 반성이라는 지적도 나온다. 더불어민주당 초선이자 20~30대 의원인 "아당·인론·청년 때문에 참패... 오영환·이소영·장경태·장철민·전용기 의원 5명은 9일 입장문을 내 "민주당 참패 원인은 저희들을 포함한 민주당의 착각과 판에 있었음을 자인한다"고 밝혔다. 이들은 국회 소통관에서 입장문..	392	

Plus + 주요 이슈별 언급량 변화 주제 관련 이슈 발생일을 기준으로 지난달 대비 총 언급량 변화(증가율)를 확인할 수 있습니다(월 기준)

April 15, 2021	대한민국 제21대 국회의원 선거	증가율 : -34.07%
----------------	-------------------	---------------

## [ 주제별 페이지 ]

### 1. 주요 시점별 여론 분석

- 시기별 여론 반응  
주요 변곡점 \*별 시계열 그래프를 통해 SNS 언급량 수치 변화(월단위)
- SNS 여론 반응  
변곡점 구간 SNS 여론 감성 분포

#### \*변곡점 선정 기준

- Point 1 : 토픽 별 영향도 높은 구간 (평균 언급량 이상인 값 중 긍정/부정 변동 구간, 주요 법률 정책 이슈 발생 주변 구간 등 고려)
- Point 2 : SNS 여론 언급량 최다 기록 구간(토픽 별 Max 값 추출)

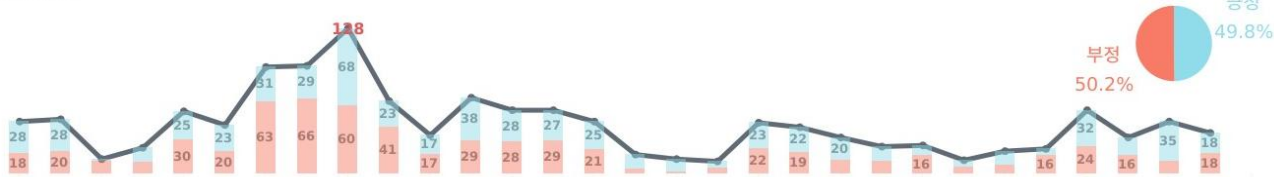
## 5-3 결과물 소개

**주요 시점별 여론 분석** 여론 이해에 필요한 주요 시점을 중심으로 시기별 여론 추이를 확인할 수 있습니다(Point 지점을 클릭 해보세요!)

청년 정책의 시기별 여론 반응



SNS 여론 반응



### AI 분석

AI 정책 제안 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

이달의 정책 주제

제안 내용

청년 구직자 지원금 제도 강화

청년들의 어려운 취업 환경을 겪는 것을 돕기 위해 구직 활동을 지원하는 지원금 제도를 강화할 수 있습니다. 지원금의 지급 규모를 확대하거나 지원 대상을 확대하는 방..

청년 노동자를 위한 개혁 정책

청년 노동자들의 근무환경 개선을 위한 규정 개혁 등 청년 노동자의 권익을 보호하고, 임금 및 기타 근로조건을 개선하는 정책을 추진할 수 있습니다.

청년 소통 센터 구축

청년들이 정치·경제·사회적 이슈에 대해 의견을 나누고, 정보를 얻을 수 있는 플랫폼을 만드는 것이 필요합니다. 이 센터에서는 청년들의 의견이 정책 결정 과정에 반영..

청년 지원 위원회의 설립

청년 문제를 전담하여 다루는 위원회의 설립하는 것이 필요합니다. 이 곳은 청년들이 직면한 문제를 집중적으로 연구하고 해결방안을 마련하는 역할을 하게 됩니다.

AI 뉴스 요약 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

댓글 수(만)

날짜	뉴스1	한겨레	댓글 수(만)
April 9, 2021	684	1,868	392

Plus + 주요 이슈별 언급량 변화 주제 관련 이슈 발생일을 기준으로 지난달 대비 총 언급량 변화(증가율)를 확인할 수 있습니다(월 기준)

April 15, 2021	대한민국 제21대 국회의원 선거	증가율 : -34.07%
----------------	-------------------	---------------

## [ 주제별 페이지 ]

### 1. 주요 시점별 여론 분석

- 시기별 여론 반응  
주요 변곡점\*별 시계열 그래프를 통해 SNS 언급량 수치 변화(월단위)
- SNS 여론 반응  
변곡점 구간 SNS 여론 감성 분포

\*변곡점 선정 기준

- Point 1 : 토픽 별 영향도 높은 구간 (평균 언급량 이상인 값 중  
긍/부정 변동 구간, 주요 법률 정책 이슈 발생 주변 구간 등 고려)
- Point 2 : SNS 여론 언급량 최다 기록 구간(토픽 별 Max 값 추출)

### 2. AI 분석(종합 페이지와 동일)

- AI 정책 제안  
여론에서 가장 중요한 주제를 뽑아 정책 아젠다 생성
- AI 지역별 뉴스 요약  
주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공



## 5-3 결과물 소개

**주요 시점별 여론 분석** 여론 이해에 필요한 주요 시점을 중심으로 시기별 여론 추이를 확인할 수 있습니다(Point 지점을 클릭 해보세요!)

청년 정책의 시기별 여론 반응



SNS 여론 반응



### AI 분석

AI 정책 제안 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

이달의 정책 주제

제안 내용

청년 구직자 지원금 제도 강화

청년들의 어려운 취업 환경을 겪는 것을 돕기 위해 구직 활동을 지원하는 지원금 제도를 강화할 수 있습니다. 지원금의 지급 규모를 확대하거나 지원 대상을 확대하는 방...

청년 노동자를 위한 개혁 정책

청년 노동자들의 근무환경 개선을 위한 규정 개혁 등 청년 노동자의 권익을 보호하고, 임금 및 기타 근로조건을 개선하는 정책을 추진할 수 있습니다.

청년 소통 센터 구축

청년들이 정치·경제·사회적 이슈에 대해 의견을 나누고, 정보를 얻을 수 있는 플랫폼을 만드는 것이 필요합니다. 이 센터에서는 청년들의 의견이 정책 결정 과정에 반영...

청년 지원 위원회의 설립

청년 문제를 전담하여 다루는 위원회를 설립하는 것이 필요합니다. 이 곳은 청년들이 직면한 문제를 집중적으로 연구하고 해결방안을 마련하는 역할을 하게 됩니다.

AI 뉴스 요약 주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공합니다

댓글 수(건)

날짜	뉴스1	한겨레	댓글 수(건)
April 9, 2021	684		
	1,868		
		392	

Plus + 주요 이슈별 언급량 변화 주제 관련 이슈 발생일을 기준으로 지난달 대비 총 언급량 변화(증가율)를 확인할 수 있습니다(월 기준)

April 15, 2021	대한민국 제21대 국회의원 선거	증가율 : -34.07%
----------------	-------------------	---------------

## [ 주제별 페이지 ]

### 1. 주요 시점별 여론 분석

- 시기별 여론 반응  
주요 변곡점\*별 시계열 그래프를 통해 SNS 언급량 수치 변화(월단위)
- SNS 여론 반응  
변곡점 구간 SNS 여론 감성 분포

\*변곡점 선정 기준

- Point 1 : 토픽 별 영향도 높은 구간 (평균 언급량 이상인 값 중  
긍/부정 변동 구간, 주요 법률 정책 이슈 발생 주변 구간 등 고려)
- Point 2 : SNS 여론 언급량 최다 기록 구간(토픽 별 Max 값 추출)

### 2. AI 분석(종합 페이지와 동일)

- AI 정책 제안  
여론에서 가장 중요한 주제를 뽑아 정책 아젠다 생성
- AI 지역별 뉴스 요약  
주요 시점의 뉴스 요약으로 여론 판단의 근거를 제공

### 3. Plus + 주요 이슈별 언급량 변화

- 주제 관련 이슈 발생일\* 기준, 전월 대비 익월의 총 언급량 변화  
및 증가율 확인

(익월 - 전월)/전월 \* 100

\*이슈 데이터 근거 - 제공된 뉴스 데이터  
(단, 법률 제정/개정 정보에 한해 '국가법령정보센터' 참고)



인공지능 기술을 활용한 대시보드를 통해  
빅데이터를 시각화하여  
의미 (여론, 정책, 법률 간 관계 이해도) 도출



개인

### 여론 반영 법안 구성

주요 시점 별 여론을 취합하여  
정책 아젠다 설정에 반영



사회

### 지역별 정책 사각지대 해소

지역별 뉴스 분석을 통해  
맞춤형 정책 분석 및  
체감도 향상



정부

### 분석 효율성 및 접근성 제고

AI 대시보드 활용 분석으로  
시간 및 비용 감소

06

Demo  
시연

# Demo

## 6-1 모델 시연

1. 패키지 설치

- 환경설정 V100 GPU, 고용량 RAM, 시스템 RAM 4GB, GPU RAM 2GB

```
In [3]: %capture
!pip install --force-reinstall numpy==1.21
!pip install torch==1.8.1
!pip install openpyxl
import openpyxl

In [4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import urllib.request
from collections import Counter #리스트 원소들의 개수 파악
from sklearn.model_selection import train_test_split

##2 데이터 불러오기

In [5]: df=pd.read_excel('../데이터/Processed_Data/1_rent_data.xlsx')

In [6]: df.head()

Out [6]:
```

Unnamed: 0	Date	User	Contents	Type	Summary	Topic
0	2022/07/01	hngc3ee0	po ??해그국장 개발방우?) werln부동산 전체 계약 전, 일상의 선택은?!\n...	트위터	극장 개 품반 부동산 전체 계약 전 선택 계 약 적전 동기부 흥 반 품공 최인 부동산...	전세사 기
1	2022/07/01	thang4u	"계약 전" 알림 (n1) 공인중개사사무소의 중 상종류 대부분 확인한다. : 공인중개...	블로그	장용 전세 전세 사기 피해 예방법 계약 전 공인 중개사 사무소 결상 등록 확...	전세사 기
2	2022/07/01	디젤메니아	답답하고 걱정스런 마음에 도움을 받을 곳이 없어 글을 남겨봅니다. \n\n\n...	커뮤니 티	오피스텔 전세 사기 예방법 어오 잘 아 시 는 분 계신가요 걱정 마음 도움 곳...	전세사 기
3	2022/07/01	dosung0502	85명에게 보증금 183억 원 받아 생겨. 건물 "두 달 추가 수사 중" \n수도...	블로그	빌라 여 채 널 후지 사기 로 사민 품민 세 오년 모한 구국 기소 명보...	전세사 기
4	2022/07/01	thang4u	중의 현지는 전세사기 예방을 널리 알리고 역 은 방법을 알려줘 했으니 \n마리 예방법...	블로그	장용 전세 전세 사기 예방 꿀 후지 전세 사 기 예방 방법 예방 수 현실 방안...	전세사 기
...	...	...	...	...	...	...
47860	2023/06/26	kkj764523	6월 26일 월요일 날씨 21°/26° 비 많이 오고 있 습니다 미세먼지 중량 초미세먼...	인스타 그램	월요일 날씨 비 미세먼지 초미세 먼지 번주 소식 장마 두 조심 뉴스 전세 사기 발생...	전세사 기
47861	2023/06/26	yangpyeunggan	<2023년 6월 26일> "하루종 들는 뉴스" >> 강남 세미원 연꽃문화제 ...	인스타 그램	하루 뉴스 위클리 미워 연꽃 문화 내달 개막 제철 행사 풍성 휴일 베품시장 등크 영팔...	전세사 기

## 주제 선택:

(All)

# M'AI REPORT

HOME

청년정책

전세사기

스토킹처벌

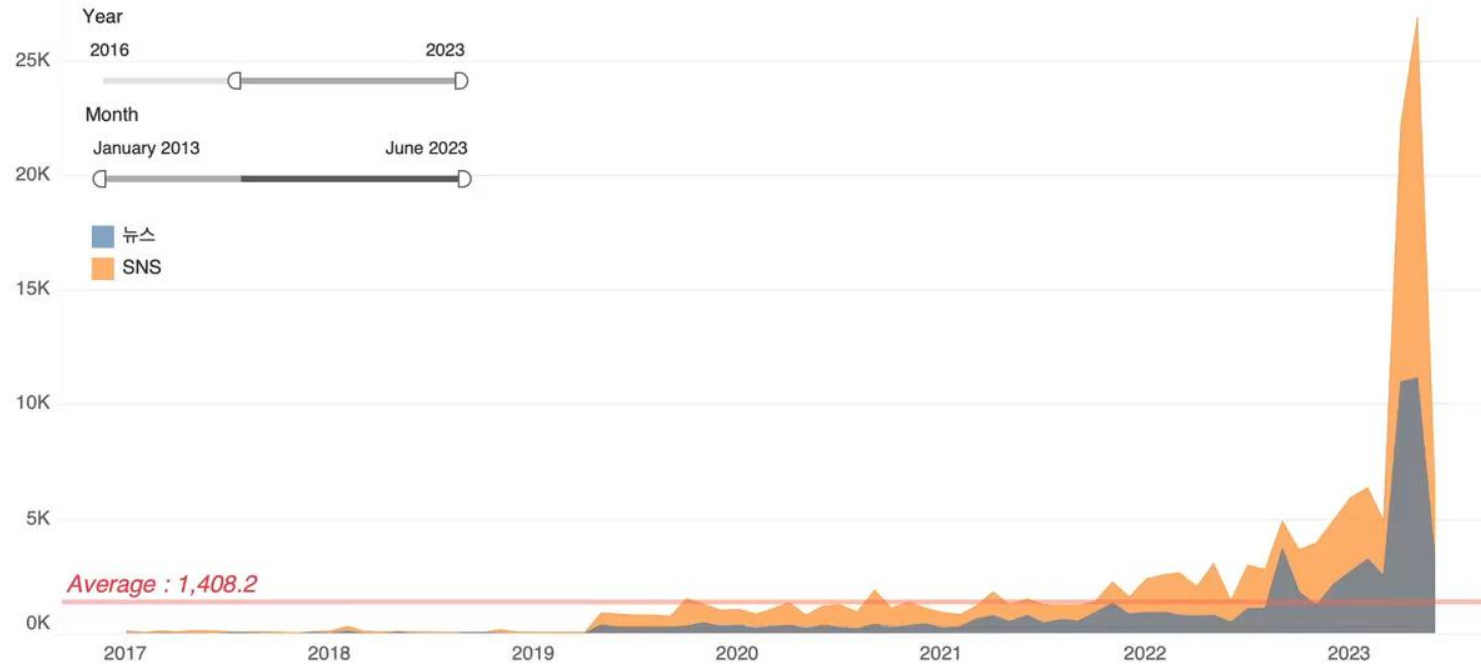
간호법



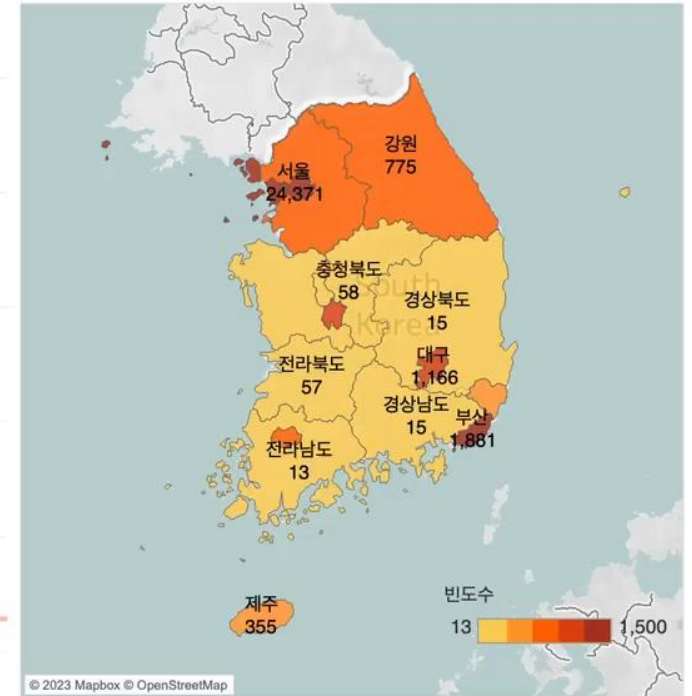
## 종합 분석

여론 언급량 추이와 연관도 높은 지역을 확인할 수 있습니다

## 시기별 SNS와 뉴스의 주제 언급량



## 지역별 뉴스 분포도



## AI 분석

**AI 정책 제안** 여론에서 가장 중요한 주제를 뽑아 정책 아젠다를 생성합니다

공간, 문화, 인구, 위원회, 조례, 행사, 소통, 개최, 계획, 기본  
 분야, 확대, 주거, 정신, 경남, 조성, 조사, 예산, 계획, 올해  
 센터, 대상, 운영, 대학, 취업, 선정, 정보, 활동, 고용, 기업  
 장애, 경제, 도시, 친화, 조직, 행정, 업무, 여성, 아동, 산업  
 후보, 국민, 문제, 의원, 대표, 정치, 선거, 위원장, 정부, 지방

1. "문화공간 개발과 활성화 정책": 이 정책은 "공간", "문화", "개최", "계획" 키워드를 중심으로 합니다. 각 지역의 문화적 특성과 역사를 고려한 멀티플렉스 문화공간을 문화공간은 인구 유입을 유도하고, 지역 활력을 촉진하면서 소통의 장을 제공하게 됩니다. 이에 대한 구체적인 계획 및 실행을 위해 '위원회'를 구성하여 '조례'를 제정

1. "선거 개혁 정책": 선택된 키워드들을 고려했을 때, 선거 과정에 대한 개혁을 추구하는 정책을 추천할 수 있습니다. 이는 후보 선시에 있어서의 공정성 확보, 국민의 있습니다. 또한 이러한 정책은 정치 현장에서 지방 선거까지 이르는 모든 수준에서 시행될 수 있도록 지향해야 합니다. 2. "국민 참여형 정책 메이킹": 정부나 문제가!

1. "주거와 예산 계획 조성" 정책: 경남 지역의 주거 문제를 살펴보고 싶습니다. 거주하는 데 필요한 시설과 인프라를 확대하여 지역 주민들의 생활 품질을 향상시키는 계획을 올해 안에 세우는 것이 중요합니다. 주거 환경의 개선은 지역 삶의 질을 높이는 데 기여하고, 정신적 안정감 또한 향상시킵니다. 2. "정신 건강 분야의 활동 확대

1. 대학 취업 정보 센터 운영 강화: 대학이 취업에 관한 정보를 제공하고, 신입생과 졸업생이 입사면접에 좀 더 잘 준비할 수 있도록 돕는 취업 정보 센터의 역할을 강화 취업 시장 동향을 파악하고 학생들에게 전달함으로써 학생들이 취업 능력을 키울 수 있도록 도와줍니다. 또한, 이런 센터들은 취업정보 제공뿐 아니라 기업들과의 네트

1. 장애인 친화적인 도시 구축 및 경제 활성화 정책: 장애인들이 일상 생활을 더 편리하게 할 수 있도록 도시의 물리적 환경을 개선해보는 것이 좋을 것입니다. 행정 업무 할 수 있습니다. 또한, 장애인의 일자리 창출을 위해 특정 산업 분야에서 장애인의 고용을 적극적으로 지원하고, 기업들에게 장애인 고용 쿼터를 시행하도록 유도하

# 출처

## ■ 참고 기사

[프레스뉴스통신]안산시, 청년지원협의체 ‘상상이상’ 회의 개최...“체감도 높은 청년정책 마련”, 2023.10.12  
<https://www.pressna.com/news/print.html?newsid=2272084>

[경기신문]“실생활에서 체감할 수 있는 인천 청년 정책 필요해”, 2023.09.24  
<https://www.kgnews.co.kr/news/article.html?no=764158>

[세계타임즈]인천 청년들이 체감할 수 있는 정책은 무엇일까?, 2023.09.22  
<https://www.thesege.com/news/view/1065574592562218>

## ■ 참고 문헌

- Jin Zhijing, Natural Language Processing for Policymaking, Handbook of Computational Social Science for Policy. 2023.

- 최준영 외 1인, 자연어처리 모델을 이용한 이커머스 데이터 기반 감성 분석 모델 구축, Journal of the Korea Convergence Society, 2020

- 남복현, <법정책의 구현형식 - 국가 법규범의 정립절차를 중심으로 ->, 법과정책연구.2016

- 최광은, <한국의 여론과 정책 연계에 관한 탐색적 연구>, 현대정치연구, 2022

## ■ 참고 자료

- kobert 이론  
<https://sktelecom.github.io/project/kobert/>

- kpfbert 이론  
<https://github.com/KPFBERT/kpfbert>

- 국가법령정보센터  
<https://www.law.go.kr/>

**감사합니다**

**Q&A**

# Appendix

1. 데이터 전처리 상세(라벨링 결과)
2. 모델별 파라미터 상세
3. 자연어 처리 상세
4. 감성 분석 모델 평가 결과
5. 대시보드 활용 분석 결과



## 1. 데이터 전처리 상세(라벨링 결과)

### Labeling

긍정/부정을 나누기 위해 정책에 대한 지지 및 홍보는 긍정(1) 나머지는 부정(0)



레이블링 데이터(5,000건)

스팸 제거(-341건)

최종 4,659건 레이블링 처리

#### \*레이블링 기준

- 긍정(1) : 주제에 대한 찬성, 지지, 홍보
- 부정(0) : 주제에 대한반대, 비판, 개선 요구



레이블링 결과 : 전체 학습 데이터 중

부정(54%) , 긍정(46%)

## 2. 모델별 파라미터 상세

환경설정: V100 GPU, 고용량 RAM, 시스템 RAM 4GB, GPU RAM 2GB

### koBERT 모델 파라미터

```
# # 하이퍼 파라미터 설정
max_len = 100
batch_size = 64
warmup_ratio = 0.1
num_epochs = 10
max_grad_norm = 1
log_interval = 200
learning_rate = 5e-5
```

### LDA 파라미터

```
import gensim
from gensim import corpora
import pandas as pd # 필요한 패키지를 import

lda_model = gensim.models.LdaModel(LdaModel(corpus=corpus,
                                             id2word=id2word,
                                             num_topics=num_topics,
                                             random_state=50,
                                             update_every=1,
                                             chunksize=10,
                                             passes=passes,
                                             alpha='symmetric',
                                             iterations=50,
                                             per_word_topics=True)

import pyLDAvis
import pyLDAvis.gensim_models as gensimvis
```

### kpfBERTsumm 모델 파라미터

```
from transformers import BertTokenizerFast, BertModel

BERT_MODEL_NAME = "jinmang2/kpfbert"
tokenizer = BertTokenizer.from_pretrained(BERT_MODEL_NAME)

MAX_TOKEN_COUNT = 512
N_EPOCHS = 5
BATCH_SIZE = 4
```

### GPT 4.0 파라미터

```
import openai

def generate_policy_recommendation(policy, keywords):
    question = f"한국의 {policy}을 강화하고 발전시키기 위한 방법을 찾고 싶습니다. LDA 모델링 결과로 언급된 키워드 중  
{','.join(keywords)}을 활용하여 어떤 정책을 추천하시겠습니까?"

    openai.api_key = "openAI_key"
    response = openai.ChatCompletion.create(
        model="gpt-4",
        messages=[
            {"role": "system", "content": "You are a helpful assistant."},
            {"role": "user", "content": question}
        ]
    )

    return response['choices'][0]['message']['content']
```

## 2. 자연어 처리 상세

### 토큰화

Mecab을 이용해 명사 추출 및 띄어쓰기 단위 구분(토큰화) 실행

### 정수 인코딩

텍스트를 숫자로 처리하도록 학습, 평가 데이터에 정수 인코딩을 수행

### 패딩

하나의 행렬로 만들기 위해 서로 다른 길이의 샘플들의 길이를 동일하게 만들

감성 분석 Model 에 적용, 평가

### 예시

#### 토큰화 전

#공예업체\_취업을 원하는 경북의 취준생들에게 꿀 정보??

#### 토큰화 후

공예 업체 취업 경북 취 준 정보

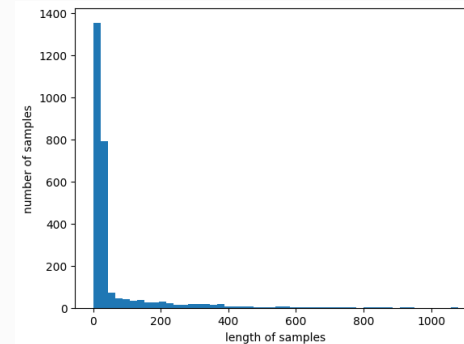
#### 정수 인코딩

{‘공예’: 1, ‘업체’: 2, ‘취업’: 3, ‘경북’: 4 ...}

단어 집합(vocabulary) 생성(단, 빈도수가 낮은 단어는 제거(빈도 수 2 이하)), 각 단어에 고유한 정수 부여

#### 패딩

[[0 0 0 0 ... 19 4]]



### 3. 감성 분석 모델 평가 결과

모델	Test Accuracy	Test Recall		Test F1-score	
		0(부정)	1(긍정)	0(부정)	1(긍정)
KoBERT (GELU)	0.824	0.889		0.889	
KoBERT & GELU ensemble lr = 0.00005	0.787	0.778		0.778	
KoBERT lr = 0.00002	0.809	0.833		0.835	
KoBERT lr = 0.00005	0.803	0.833		0.833	
BiLSTM & LSTM ensemble	0.79	-		-	
LSTM	0.76	0.85	0.68	0.78	0.74
BiLSTM	0.77 (이전 0.74)	0.85 (이전 0.78)	0.69 (이전 0.70)	0.80 (이전 0.75)	0.74 (이전 0.73)
GRU	0.74	0.68	0.79	0.72	0.76

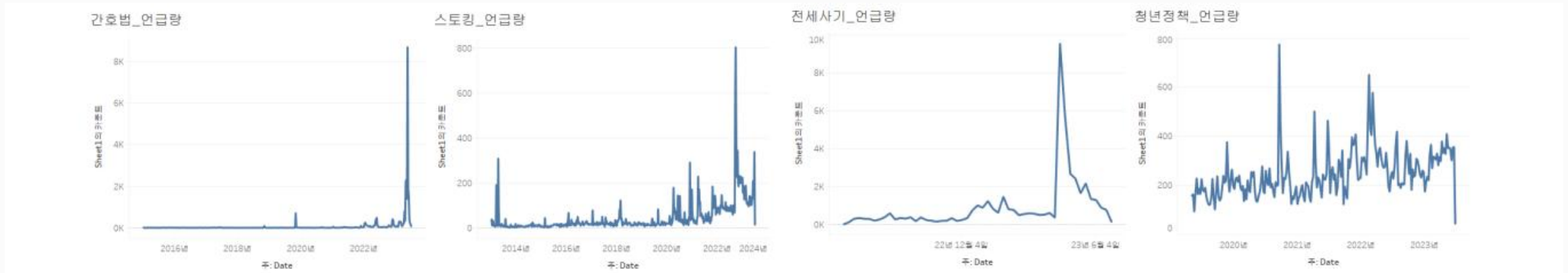
## 4-1. 대시보드 활용 분석 결과

### 1. 전체 여론 흐름(토픽 별 변곡점)

- 모든 토픽의 여론 형성 트렌드의 공통점이 있을까?

- 분석 결과 : 4가지 토픽 모두 사건 등 이슈의 발생 후 여론 증가 추세

- 간호법, 청년정책: 대선, 총선 등 정책의 내용에 영향을 미치는 이슈 발생 후 여론 언급량 증가 추세
- 스톡킹 처벌, 전세사기: 피해 사례 등 범죄 이슈 발생 후 여론 증가 추세



구분	이슈 시기	이슈 내용	유형	결과
간호법	2023/05	윤대통령 간호법 제정 거부권 행사	정책 - 제정 결렬	언급량 최다 기록
	2022/01	대선 후보 간호법 제정 공약	사회 - 대선	해당 변곡점 이후 여론 증가 추세
스토킹 처벌	2022/09	신당역 역무원 스토킹 살해	사건 - 범죄	언급량 최다 기록
	2020/04	황교안 'n번방 호기심' 발언 비난	사회 - 국회의원 발언 논란	해당 변곡점 이후 여론 증가 추세
전세사기	2023/04	세입자 전세금 125억원 가로챈 '건축왕' 피해자 숨져	사건 - 피해자 사망	언급량 최다 기록
	2023/01	빌라왕 전세사기 피해사례 / 국토부, 피해자 구제안 마련	사건 - 피해	해당 변곡점 이후 여론 증가 추세
청년정책	2022/03	대선(20대 대통령선거 - 투표 연령이 2020년 만 18세로 낮춰진 뒤 이뤄지는 첫 대선)	사회 - 대선	언급량 최다 기록
	2021/04	총선	사회 - 총선	해당 변곡점 이후 여론 증가 추세

## 4-2. 대시보드 활용 분석 결과

### 2. 정책 관련 이슈 언급량 영향도

- 가설 : 정책 관련 내용이 여론 형성에 미치는 영향이 있는가?
- 결과 : 간호법>청년정책>스토킹처벌>전세사기 순으로 이슈 보도 이후 여론 증가(월 기준)
  - 4개 토픽 총 평균 약 28% 증가

Topic	Date	issue	D-30	D-day	D+30	증가율
간호법	2021-03-25	00:00:00간호법 법안 발의	1016	3	1906	87.59843
간호법	2023-02-09	00:00:00패스트트랙(신속처리안건 지정) 상정	6668	109	5578	-16.3467
간호법	2023-03-23	00:00:00국회 본회의 부의	4870	44	14670	201.232
간호법	2023-04-27	00:00:00국회 본회의 통과	18488	677	27500	48.74513
간호법	2023-05-16	00:00:00윤대통령 간호법 거부권 행사	30192	2663	17390	-42.402
스토킹처벌	2021-04-20	00:00:00스토킹처벌법 제정	2003	31	1202	-39.99
스토킹처벌	2021-10-21	00:00:00스토킹처벌법 시행	1014	75	1947	92.01183
전세사기	2023-02-28	00:00:00인천 건축왕 사태 피해자 사망	6833	65	4770	-30.1917
전세사기	2023-06-01	00:00:00전세사기피해자법 제정	25960	234	6302	-75.7242
청년정책	2020-02-04	00:00:00청년기본법 제정	1097	40	848	-22.6983
청년정책	2020-12-23	00:00:00제1차 청년정책 기본계획('21 ~ '25년)	1205	59	969	-19.5851
청년정책	2021-08-17	00:00:00청년기본법 일부개정(청년의 정책 참여 확대)	1092	37	1474	34.98168
청년정책	2023-03-21	00:00:00청년기본법 일부개정(취약계층 청년 권익 강화)	4974	49	12226	145.7982

평균 27.95% 증가율