

# 포트폴리오

---

이은호

GitHub: [github.com/eunnnholee/Portfolio](https://github.com/eunnnholee/Portfolio)

Blog: [blog.naver.com/eunnho\\_lee](https://blog.naver.com/eunnho_lee)

# 목차

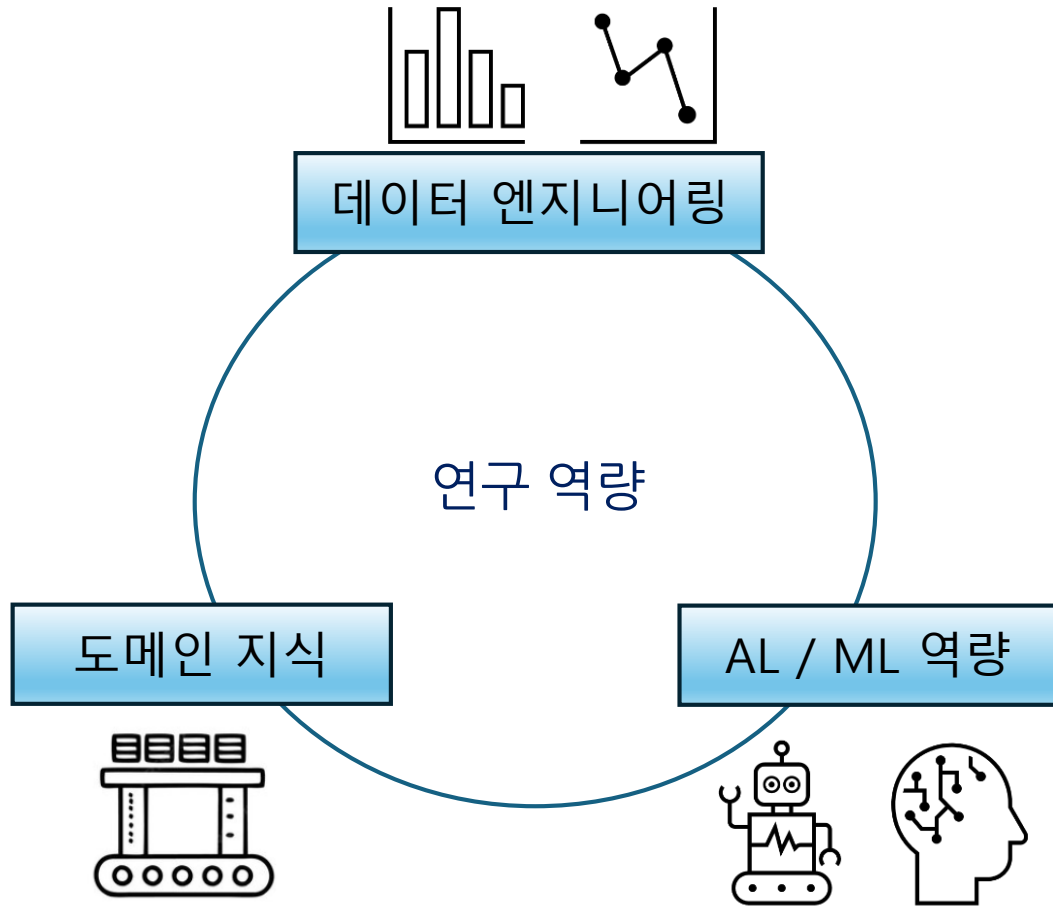
---

## 1. 프로젝트 소개

- A. LG AI 해커톤 : 디스플레이 제조 공정을 위한 결함 예측 AI 모델
- B. 데이터마이닝 프로젝트 : 육아지원정책: KoBERTopic과 SNA를 통한 심층 분석
- C. 딥러닝 프로젝트: 시각장애인을 위한 보조 기능 - Image Captioning
- D. 동서발전 공모전: 기상데이터 기반 풍력 발전 예측량 예측

## 2. 결론

## 프로젝트 소개



### 기상데이터 기반 풍력 발전량 예측

- 성과 : 본선 진출 (6등 이내)
- 시기 : 2024.10 - 2024.11
- 역할 : 시계열 데이터 분석 및 딥러닝 모델 구현

### 디스플레이 제조 공정을 위한 결함 예측 AI 모델

- 성과 : 36등 / 740팀
- 시기 : 2024.07 - 2024.08
- 역할 : 데이터불균형 해소 및 파생변수

### 시각장애인을 위한 보조 기능 - Image Captioning

- 목표 : 시각장애인을 위한 보조 기능으로서의 이미지 캡셔닝 기술모델 비교 및 효용성 평가
- 시기 : 2023.10 - 2023.11
- 역할 : InceptionV3 구조 제안 및 모델 설계

### 육아지원정책: KoBERTopic과 SNA를 통한 심층 분석

- 목표 : 산발된 육아휴직 관련 약 4,000개 텍스트 데이터 7개의 주요 질문으로 재구성
- 시기 : 2023.03 - 2023.07
- 역할 : 크롤링 및 텍스트마이닝, SNA 방법론 제안 및 Node와 Edge 설계

A. 프로젝트 소개

기상데이터 기반 풍력 발전량 예측  
동서 발전 공모전

❖ Raw Dataset : LDAPS, SCADA, Target(풍력발전량)

	dt	elevation	land_cover	surf_rough	frictional_vmax_50m	frictional_vmin_50m	pressure	relative_humid	specific_humid	temp_air
0	2020-01-02 00:00:00+09:00	0.15625	0.021042	0.004259	3.177766	2.558847	103028.843750	61.177334	0.003413	280.260773
1	2020-01-02 01:00:00+09:00	0.15625	0.021042	0.004250	3.907736	3.248139	103004.507812	61.698036	0.003413	280.260681
2	2020-01-02 02:00:00+09:00	0.15625	0.021042	0.004248	3.552126	3.160641	102999.546875	62.101166	0.003418	280.268921
3	2020-01-02 03:00:00+09:00	0.15625	0.021042	0.004248	3.269210	2.882998	102988.968750	61.944283	0.003418	280.276672
4	2020-01-02 04:00:00+09:00	0.15625	0.021042	0.004248	3.065275	2.887952	102937.523438	61.080189	0.003418	280.266052
...	...	...	...	...	...	...	...	...	...	...
917065	2022-12-31 19:00:00+09:00	14.84375	1.000000	0.200022	4.003626	3.589689	102833.937500	64.626816	0.003016	274.376099
917066	2022-12-31 20:00:00+09:00	14.84375	1.000000	0.200022	3.916099	3.548627	102825.851562	66.772079	0.003124	273.657501
917067	2022-12-31 21:00:00+09:00	14.84375	1.000000	0.200022	4.573833	3.845803	102831.773438	68.919631	0.003174	273.419952
917068	2022-12-31 22:00:00+09:00	14.84375	1.000000	0.200022	5.136635	4.590230	102843.687500	67.002106	0.003124	273.396545
917069	2022-12-31 23:00:00+09:00	14.84375	1.000000	0.200022	5.615887	5.049034	102829.242188	69.130402	0.003174	273.379211

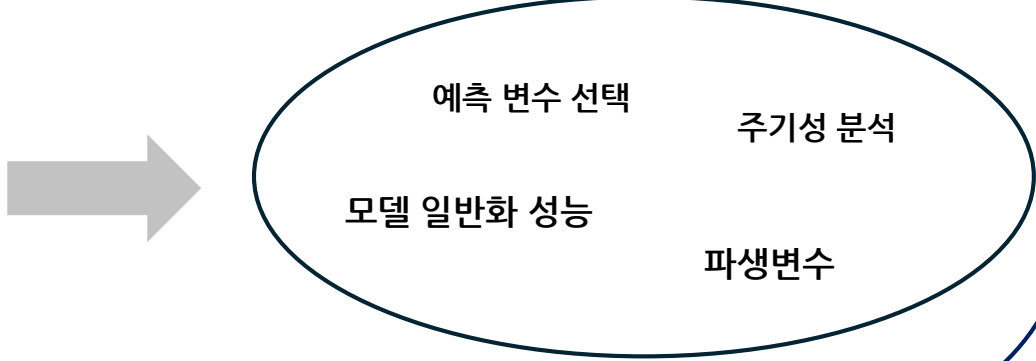
917070 rows x 15 columns

Date/Time	WTG Name	WTG Serial	Availability Forecast	Availability Requested	Availability Scheduled	Availability Technical	Energy Production Active Power (MW)	Generator Average Temp. (°C)	Grid Active Power (MW)	Grid Reactive Power (MW)	Hydraulic System Pressure (bar)	Nozzle Air Density (kg/m³)	Nozzle Outlet Temp. (°C)	Nozzle Outlet Position (deg)	Nozzle Wind Direction (deg)	Nozzle Wind Speed (m/s)	Rotor Speed (rpm)	Yaw cable windup position (deg)	Yaw cable windup position (deg)	Availability Full Performance	Rotor Blade 1 Pos. (deg)	Rotor Blade 2 Pos. (deg)	Rotor Blade 3 Pos. (deg)	Rotor Motor 1 Pos. (deg)	Rotor Motor 2 Pos. (deg)	Rotor Pitch 1 Angle (deg)	Rotor Pitch 2 Angle (deg)	Rotor Pitch 3 Angle (deg)	
2020-01-01 00:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	114.4763	114.4763	4.34366	10.88353	1.288166	-2.65704	11.88801	110.1658	5.357079	9.047179	683.8388	683.8388	0	0.1018	-0.12633	-0.07177	0	0	0	0	0	0
2020-01-01 01:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	115.7154	115.7154	75.1993	113.36	1.08868	-2.80714	11.08801	110.1658	5.403822	9.915423	683.8388	683.8388	0	0.10664	-0.12343	-0.07094	0	0	0	0	0	0
2020-01-01 02:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	114.4763	114.4763	34.7059	10.93101	1.293474	-2.80434	11.25154	114.3375	5.391346	9.681194	687.7111	687.7111	0	0.10173	-0.12873	-0.01487	0	0	0	0	0	0
2020-01-01 03:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	114.4763	115.7154	25.9508	10.88353	1.2512	-2.87688	10.77036	104.4794	4.40931	8.54916	679.1621	679.1621	0	0.086267	-0.1399	-0.02773	0	0	0	0	0	0
2020-01-01 04:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	117.9415	120.3448	72.2055	130.351	1.289939	-2.91383	10.79654	112.9756	5.547621	9.884538	681.7849	681.7849	0	0.107133	-0.12587	-0.01733	0	0	0	0	0	0
2020-01-01 05:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	117.7044	120.3906	78.13257	130.601	1.28826	-2.84899	12.7974	110.3906	4.910542	9.793933	684.799	684.799	0	0.096133	-0.13313	-0.02463	0	0	0	0	0	0
2020-01-01 06:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	121.2664	121.2664	81.278	130.988	1.28978	-2.86541	11.2241	111.2241	5.529161	10.03641	682.1395	682.1395	0	0.111267	-0.12127	-0.01153	0	0	0	0	0	0
2020-01-01 07:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	118.2664	118.2664	54.9571	128.963	1.28759	-2.97031	11.5135	114.3375	6.049042	10.7429	685.9106	685.9106	0	0.111467	-0.11713	-0.01847	0	0	0	0	0	0
2020-01-01 08:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	117.4484	117.4484	85.0176	130.009	1.2842	-2.89537	11.6817	111.507	5.387283	10.77442	690.1186	690.1186	0	0.114687	-0.11517	-0.01927	0	0	0	0	0	0
2020-01-01 09:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	123.9376	123.9376	47.5811	132.482	1.28935	-2.9444	14.2918	115.0024	5.287907	9.82527	686.2912	686.2912	0	0.108333	-0.12403	-0.01233	0	0	0	0	0	0
2020-01-01 10:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	121.1871	121.1871	26.371	85.378	1.28978	-2.9862	11.623	104.4794	4.40931	8.54916	679.1621	679.1621	0	0.11195	-0.10777	-0.01913	0.002647	0.002647	0.002647	0.002647	0.002647	
2020-01-01 11:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	120.3448	118.2664	66.6105	140.53	1.27867	-2.96641	10.77442	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 12:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 13:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 14:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 15:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 16:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 17:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 18:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 19:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 20:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 21:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 22:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-01 23:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 00:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 01:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 02:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 03:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 04:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 05:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 06:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 07:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 08:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 09:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 10:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 11:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 12:00:00+09:00	WTG001	1013-001	0	0	0	0	0.000000	116.9868	116.9868	82.3983	128.964	1.28964	-2.9524	11.2664	111.507	5.387283	10.7429	685.9106	685.9106	0	0.111267	-0.11189	-0.07763	0	0	0	0	0	0
2020-01-02 13:00:00+09:00	WTG001	1013-001	0	0	0	0	0.																						

	plant_name	end_datetime	period_hours	energy_kwh
0	경주풍력	2020-01-01T01:00:00+09:00	1	9767.578125
1	경주풍력	2020-01-01T02:00:00+09:00	1	5381.835938
2	경주풍력	2020-01-01T03:00:00+09:00	1	3021.484375
3	경주풍력	2020-01-01T04:00:00+09:00	1	4400.390625
4	경주풍력	2020-01-01T05:00:00+09:00	1	4501.953125
...	...	...	...	...
52603	영광풍력	2022-12-31T20:00:00+09:00	1	1708.529000
52604	영광풍력	2022-12-31T21:00:00+09:00	1	67.645000
52605	영광풍력	2022-12-31T22:00:00+09:00	1	0.000000
52606	영광풍력	2022-12-31T23:00:00+09:00	1	0.000000
52607	영광풍력	2023-01-01T00:00:00+09:00	1	443.374000

52608 rows x 4 columns

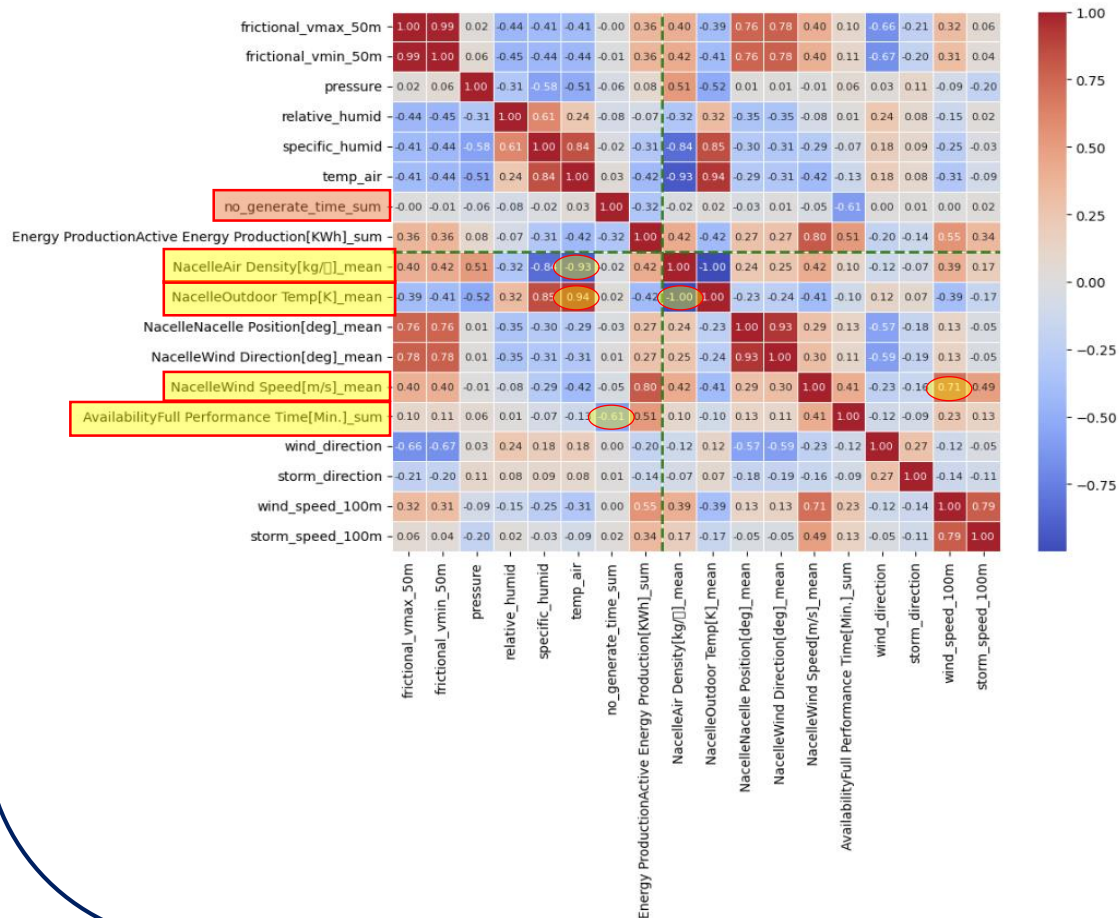
- LDAPS, SCADA 데이터 세트 : 2개
- 영광, 경주 발전 단지 관련 데이터 : 2개
  - 총 4개 데이터 세트
- 각 데이터 세트는 터빈으로 구분되어 있음
- SCADA 데이터는 학습의 Input으로 사용 불가 제약 존재(raw한 형태로 사용 불가) : 가용 필수
- LDAPS 데이터를 활용해 SCADA 데이터를 예측하는 SubModel 구축



## A. 프로젝트 소개

### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

❖ SCADA 데이터 중 어떤 변수를 예측에 활용할까? WHAT?



### Energy Production 변수와의 상관관계

1등: NacelleWind Speed

2등: AvailabilityFull Performance Time

3등: NacelleAir Density, NacelleOutdoor

4등: no\_generate\_time

최종 예측 변수 선정: no\_generate\_time

Energy Production: SCADA 데이터에 주어진 공급량으로 풍력발전량과 비례 관계  
Energy Production과 나머지 변수의 상관 관계를 분석해 최종 예측 변수를 선정

Energy Production과 상관관계가 높은 SCADA 변수를 SubModel로 예측 수행 계획  
But 1등~3등까지의 변수는 LDAPS 데이터와도 충분히 높은 상관 관계를 가짐  
LDAPS 변수로도 충분한 설명력을 갖고 있다고 판단

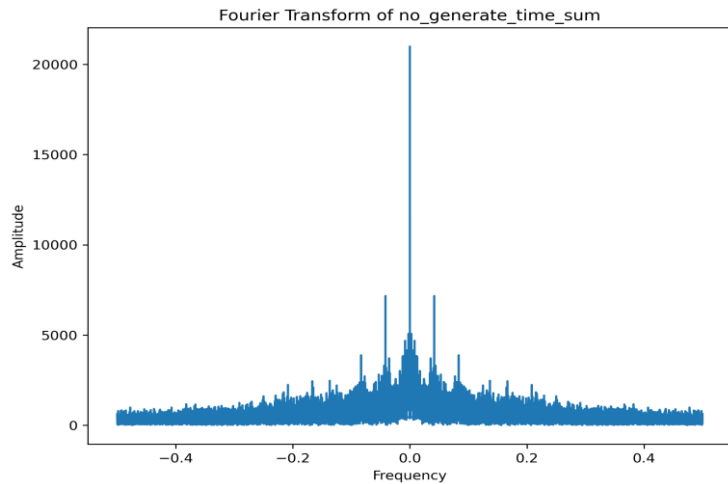
최종 예측 변수 no\_generate\_time 으로 선정

## A. 프로젝트 소개

### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

❖ no\_generate\_time 변수 어떻게 예측할까? HOW?

➤ 주기성(시계열성) 분석



Top 5 frequency peaks for no\_generate\_time\_sum:

Frequency: -0.00247, Amplitude: 5071.39006

Frequency: 0.00247, Amplitude: 5071.39006

Frequency: -0.04167, Amplitude: 7167.36434

Frequency: 0.04167, Amplitude: 7167.36434

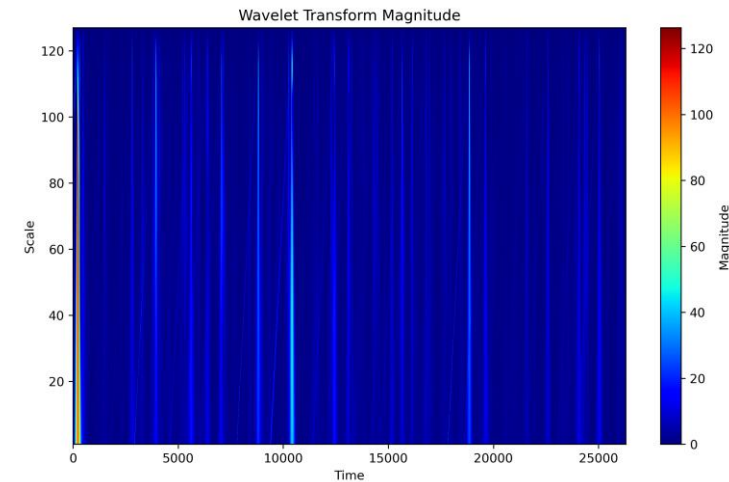
Frequency: 0.00000, Amplitude: 20977.56667

$$\text{Period} = 1 / \text{Frequency}$$

$$\text{Frequency} = 0.00247 = 404$$

$$\text{Frequency} = 0.04167 = 24$$

➤ 명확한 주기성 파악 불가



➤ 명확한 주기성 파악 불가

## A. 프로젝트 소개

### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

#### ❖ no\_generate\_time 변수 어떻게 예측할까? HOW?

##### ➤ 주기성(시계열성) 분석

#### Ljung Box Test

시계열 데이터가 유의미한 자기상관성을 가지고 있는지를 확인하는 통계적 검정

여러 개의 시차(lag)에 대해 자기상관이 동시에 0인지를 검정

추세성분과 주기성 성분을 모두 포함함

귀무가설: 데이터가 백색잡음이다.

대립가설: 데이터에 유의미한 자기상관이 있다.(백색잡음이 아니다.)

#####

no\_generate\_time\_sum

period:24, seasonal:7, trend:9697

#####

1.주기 성분의 분산 비율 (Seasonal Ratio): 0.2286

2.주기성 강도 지표 (Seasonal Strength): 0.2294

3.Ljung-Box 검정 resid manual p-value: 0.0000

3.Ljung-Box 검정 seasonal manual p-value: 0.0000

#####

no\_generate\_time\_sum

period:404, seasonal:2401, trend:17521

#####

1.주기 성분의 분산 비율 (Seasonal Ratio): 0.0297

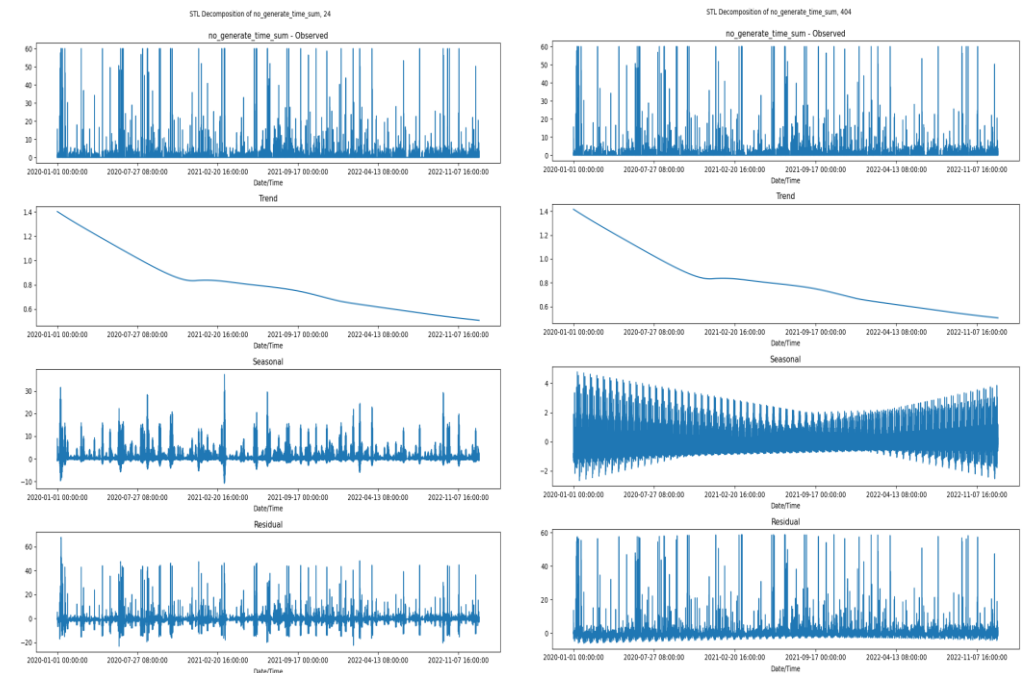
2.주기성 강도 지표 (Seasonal Strength): 0.0298

3.Ljung-Box 검정 resid manual p-value: 0.0000

3.Ljung-Box 검정 seasonal manual p-value: 0.0000

➤ 귀무가설 기각: 자기상관성 존재

#### STL Analysis



➤ 장기 감소 추세(Trend)

➤ 약한 주기성 확인

“ no\_generate\_time ” 변수의 주기성 약하게 존재

## A. 프로젝트 소개

기상데이터 기반 풍력 발전량 예측  
동서 발전 공모전

❖ no\_generate\_time 변수 어떻게 예측할까? HOW?

➤ 시계열 모델 예측

ML: SARIMAX, 프로핏

DL: CNN, LSTM, MLP

위 모델 적층 식 조합

VS

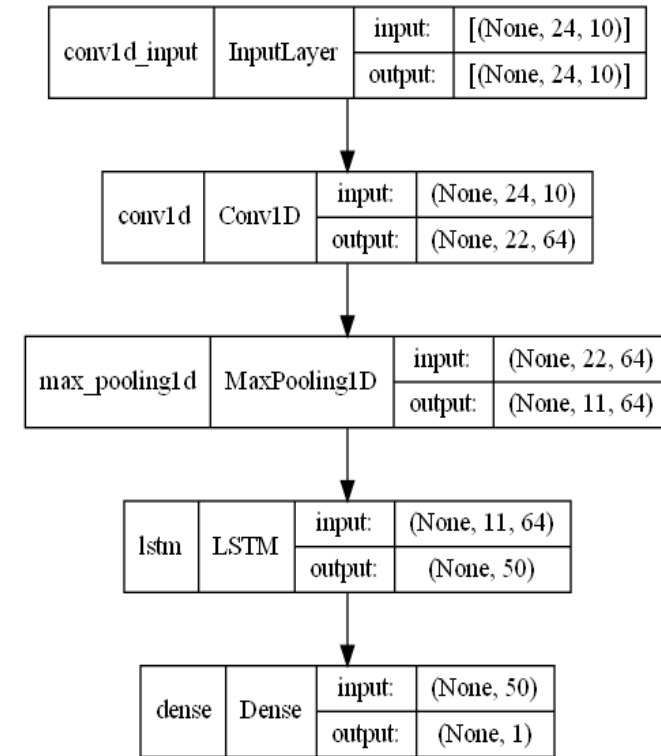


Figure 1. CNN-LSTM MODEL

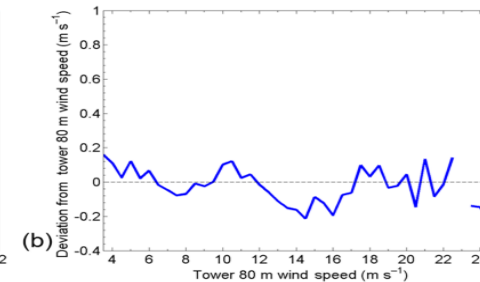
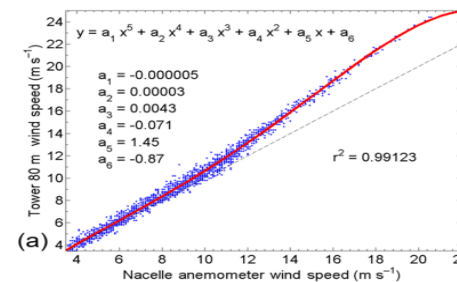
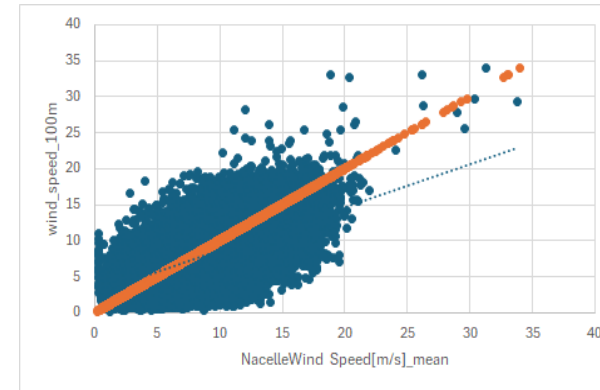
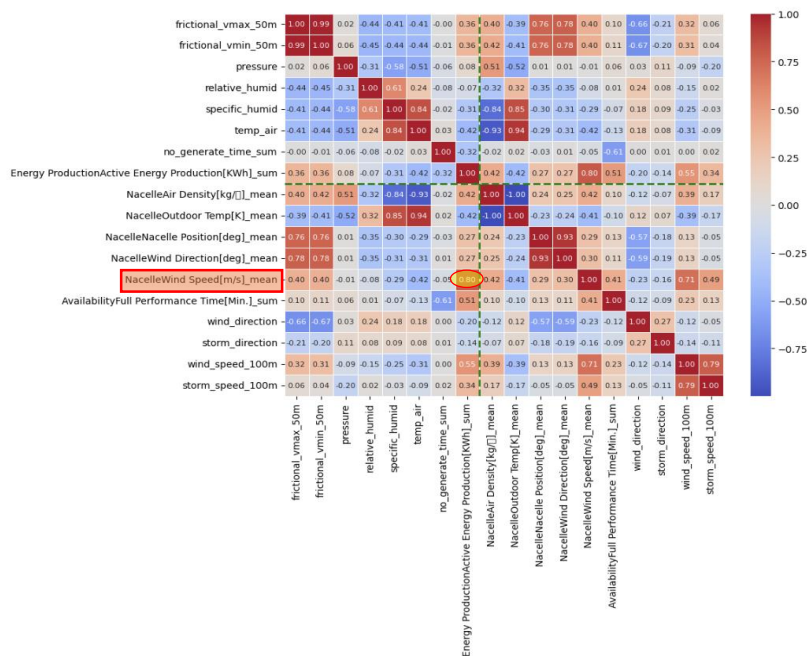
여러 모델을 테스트하고 모델 최적화를 수행함에도 모델 성능이 최소 목표치(NMAE 0.8 이하) 이상으로 나오지 않아 no\_generate\_time 예측 무리



# A. 프로젝트 소개

## 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

### ❖ SCADA 데이터 중 어떤 변수를 예측에 활용할까? WHAT?



LDAPS 데이터는 실제 데이터가 아닌, 예측된 데이터

불안정한 예측치를 갖고 있는 LDAPS의 풍속 데이터를 보완하고자 하는 목적

### NacelleWind Speed 변수를 예측한 이유

1. Energy Production 변수와의 상관 관계 가장 큼
2. LDAPS 데이터의 wind\_speed\_100m 변수의 예측 결과가 좋지 않음

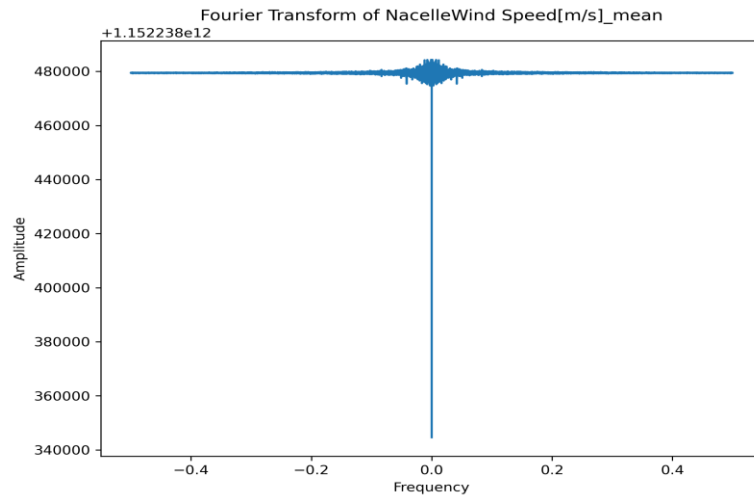
SCADA 데이터에서 가져올 변수 : **NacelleWind Speed**

## A. 프로젝트 소개

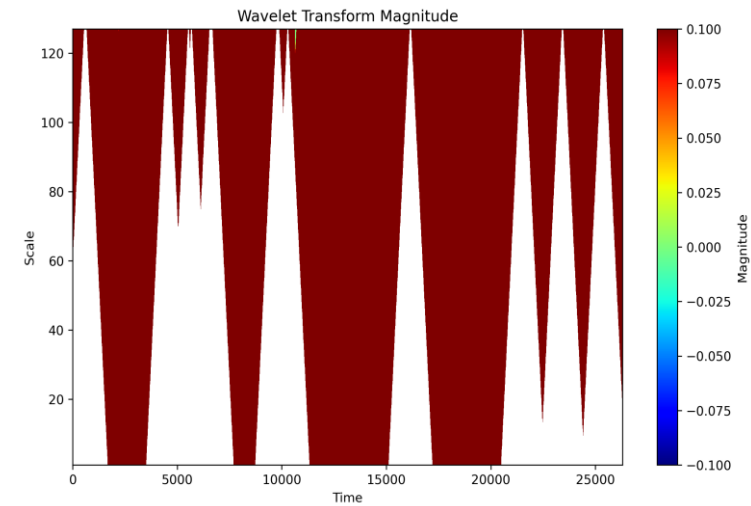
### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

#### ❖ NacelleWind Speed 변수 어떻게 예측할까? HOW?

##### ➤ 주기성(시계열성) 분석



➤ 명확한 주기성 파악 불가



➤ 패턴으로 보이나 동일한 패턴으로 보기 어려움

➤ 명확한 주기성 파악 불가

#### NacelleWind Speed

- 주기성 없다고 판단
- 시계열 모델 사용하지 않고 ML 모델을 활용하여 SubModel 생성

## A. 프로젝트 소개

### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

#### ❖ Feature Selection & Preprocessing

##### ➤ VIF 분석 & 도메인 지식

	Feature	VIFscore
0	frictional_vmin_50m	2.093306
1	specific_humid	3.301010
2	no_generate_time_sum	1.641181
3	Energy ProductionActive Energy Production[KWh]...	3.844921
4	AvailabilityFull Performance Time[Min.]_sum	6.832483
5	wind_direction	5.734714
6	storm_direction	4.150131
7	wind_speed_100m	7.512765

VIF & 도메인 지식을 활용해 변수 최종 선택

##### ➤ 각 데이터 전처리 및 대표 터빈 생성



LDAPS

wind\_speed\_100m 계산

$\text{wind\_speed\_10m} = \sqrt{[\text{'wind\_u\_10m'}]**2 + [\text{'wind\_v\_10m'}]**2}$

$\text{wind\_speed\_100m} = \text{Logarithmic\_profile}([\text{'wind\_speed\_10m'}], 10, 100, [\text{'surf\_rough'}])$

wind(storm) direction: cos, sin 변환



SCADA

10분 단위 -> 1시간 단위 (변환)

no\_generate\_time 파생변수 생성

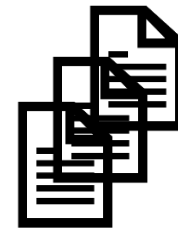
$\text{no\_generate\_time} = (\text{AvailabilityForced Outage Time} + \text{AvailabilityRequest Shutdown Time} + \text{AvailabilityScheduled Maintenance Time} + \text{AvailabilityTechnical Standby Time})$



공통

고유값 한 개 칼럼 제거

Null 값 제거 (행 기준)



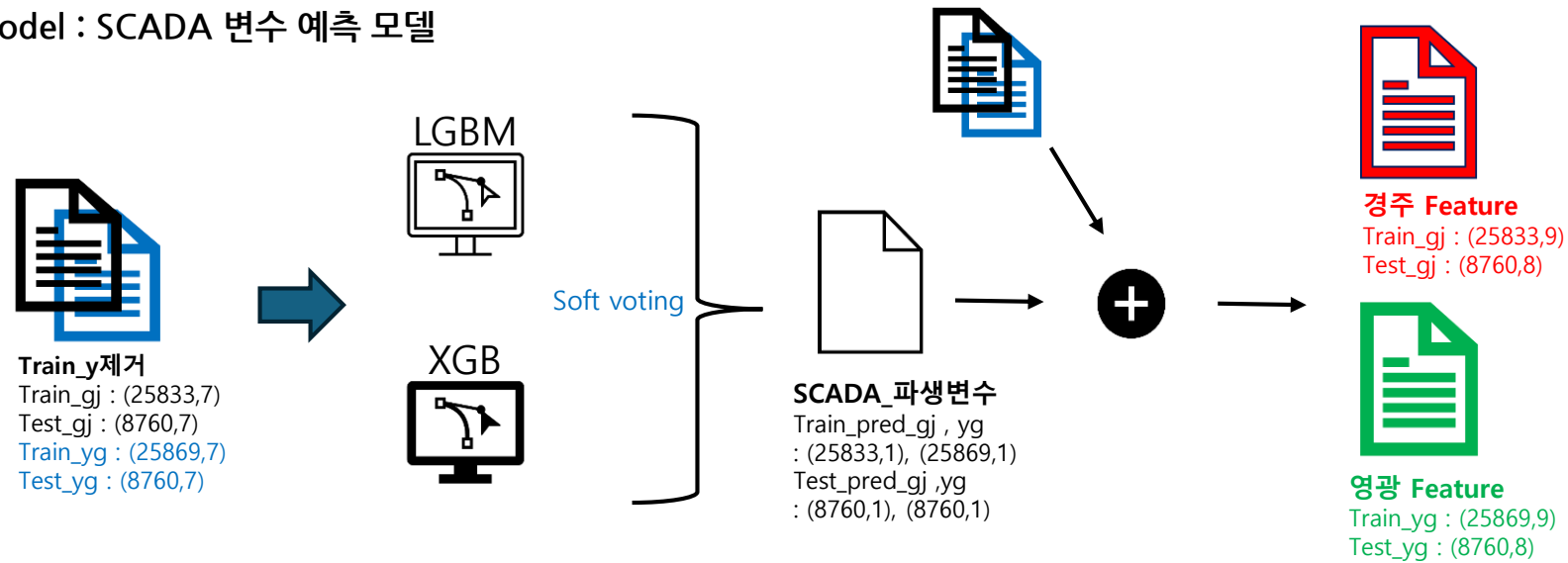
대표 터빈 생성

Date/Time 칼럼 기준으로 평균화

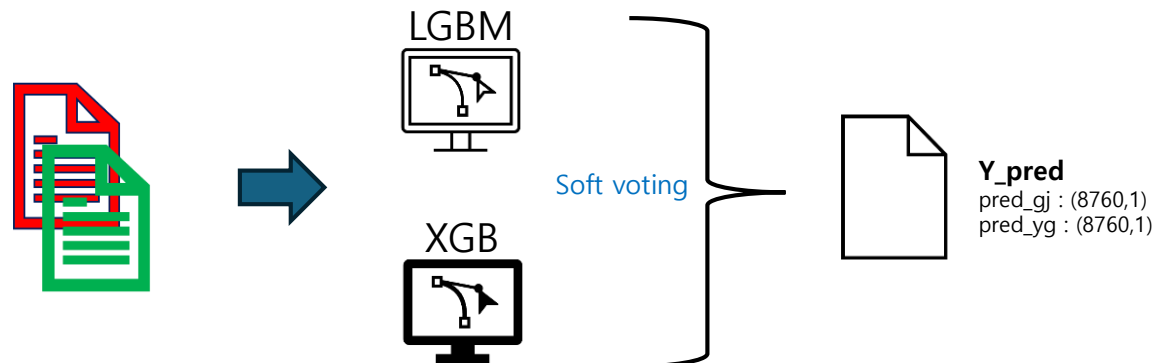
## A. 프로젝트 소개

### 기상데이터 기반 풍력 발전량 예측 동서 발전 공모전

#### 1. Sub Model : SCADA 변수 예측 모델



#### 2. Main Model : 풍력 발전량 예측 모델



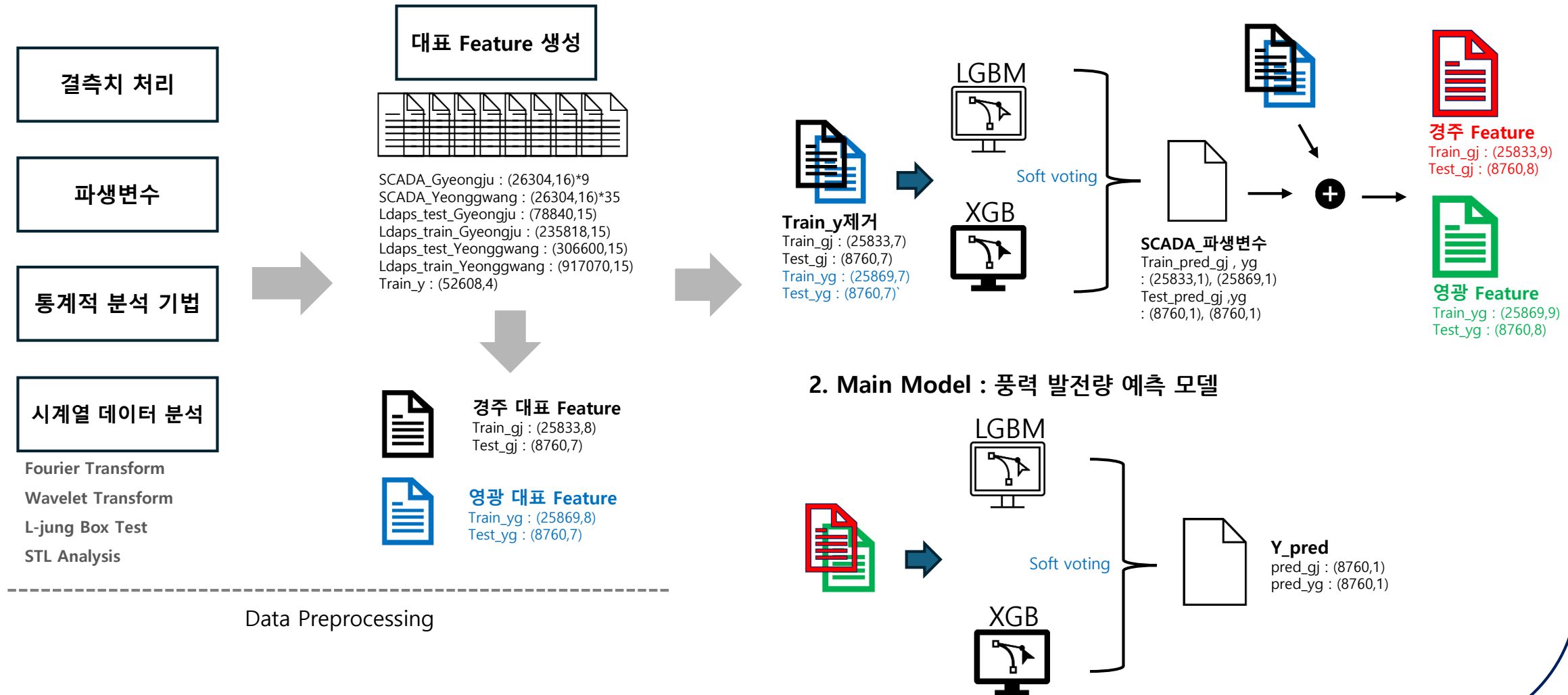
#### 최종 평가 지표

**영광 발전 단지**  
MAE: 4475.1981  
MSE: 51396541.7602  
RMSE: 7169.1381  
r2: 0.8939  
nmae: 0.0562

**경주 발전 단지**  
MAE: 1814.8325  
MSE: 6392991.4049  
RMSE: 2528.4366  
r2: 0.8228  
nmae: 0.0877

## A. 프로젝트 소개

기상데이터 기반 풍력 발전량 예측  
동서 발전 공모전



# A. 프로젝트 소개

디스플레이 제조 공정을 위한 결함 예측 AI 모델  
LG AI 해커톤

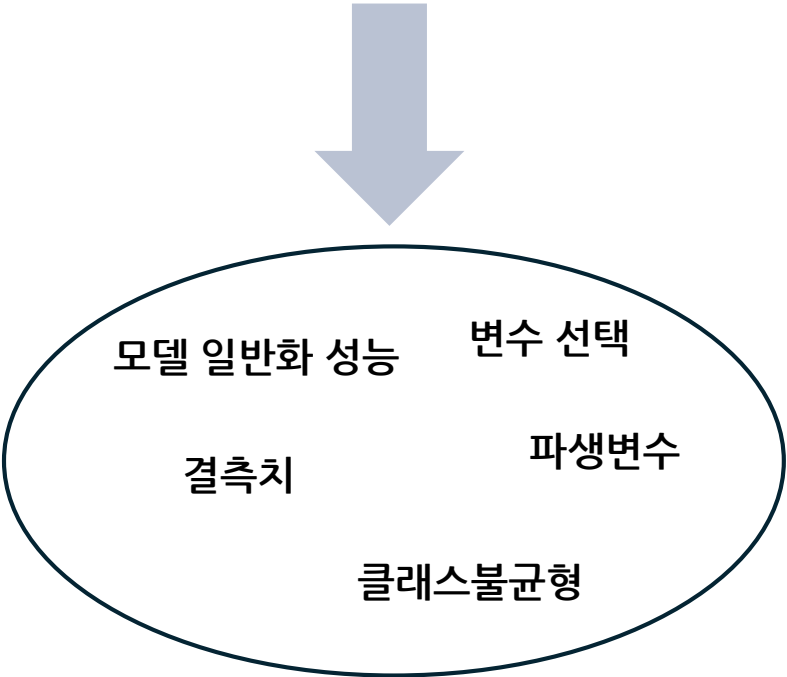
## ❖ 데이터 셋 구성



## Raw Dataset

40501	IVI-OB6	Dam Dispenser	Dam dispenser #1	AJX75334501	3J1XF434-2	1	OK	240.0	NaN	NaN	2.5
40502	IVI-OB6	Dam Dispenser	Dam dispenser #2	AJX75334501	4E1XC796-1	1	OK	1000.0	NaN	NaN	12.5
40503	IVI-OB6	Dam Dispenser	Dam dispenser #1	AJX75334501	4C1XD438-1	1	OK	240.0	NaN	NaN	2.5
40504	IVI-OB6	Dam Dispenser	Dam dispenser #2	AJX75334501	3I1XA258-1	1	OK	1000.0	NaN	NaN	12.5
40505	IVI-OB6	Dam Dispenser	Dam dispenser #1	AJX75334501	3G1XA501-1	1	OK	240.0	NaN	NaN	2.5

40506 rows x 464 columns



## A. 프로젝트 소개

디스플레이 제조 공정을 위한 결함 예측 AI 모델

LG AI 해커톤

### ❖ 구조도

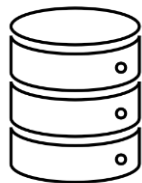
결측치 처리

파생변수

통계적 분석 기법

Chi-Square  
Anderson-Darling  
Kolmogorov-Smirnov  
Mann-Whitney U

Data Preprocessing



Train Dataset shape : (40506, 213)  
Normal Dataset shape: (38156, 213)  
AbNormal Dataset shape : (2350, 213)

Normal 0.941984  
AbNormal 0.058016  
Name: target, dtype: float64  
Class Imbalance : (16 : 1)

Class Imbalance (16:1)  
Dataset

Normal Data 비복원 추출

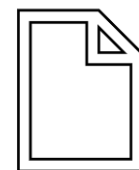
표본데이터세트 :  
6\*Normal + 1\*Abnormal



표본 데이터세트 : (14100, 213)  
(2350, 213)

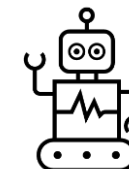


표본 데이터세트 : (14100, 213)  
(2350, 213)

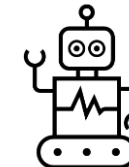


표본 데이터세트 : (9956, 213)  
(2350, 213)

Data Split (6:1)  
클래스 불균형 완화



CatBoost Model



LGBM Model

두 개의 모델의 예측결과 Soft Voting  
클래스불균형 해소 하이퍼파라미터 설정

## A. 프로젝트 소개

디스플레이 제조 공정을 위한 결함 예측 AI 모델

LG AI 해커톤

### ❖ 파생변수

#### 1. 로봇팔 좌표 : 유사한 고유값 그룹화 및 평균값 대체

목표: train\_data, test\_data의 특정 열에서 고유값을 그룹화하여 범위를 기준으로 매핑 값 생성

주요 논리:

- 열 목록을 순회하며, 고유값들을 동적으로 그룹화하여 값의 매핑 생성
- 합집합(np.union1d)으로 고유값 통합
- 값 범위( $\pm 5$ )를 기준으로 그룹핑 및 평균값 계산
- 그룹화된 값을 새로운 매핑으로 변환

생성 이유:

공정과정에서 로봇팔로 레진을 도포하는 과정 존재  
로봇팔의 좌표값(x, y, z)이 변수로 주어짐

특정 좌표값의 고유값이 유사한 값을 띄는게 다수 존재

예를 들면, 100.1, 100.2, 99.9와 같이 하나의 값 주변에 유사한 고유값이 몰려 있다는 것을 파악  
이 값은 **설비 세팅 값에서 오는 오류**라고 판단했으며, 오차 범위( $\pm 5$ )를 기준으로 그룹화 진행

흐름:

1. 초기화
  - target\_column 리스트 정리: HEDA NORMAL COORDINATE X(Y, Z) AXISStage1(2, 3) Collect Result\_Dam(Fill1, Fill2) ...
  - Unique\_values: train\_data, test\_data의 고유값 합집합으로 설정
2. 고유값 그룹화
  - 고유값이 특정 범위( $\pm 5$ )에 속하면 그룹에 추가
  - 그룹에 추가될 때 평균값 업데이트
3. 매핑 생성
  - 그룹별 매핑 딕셔너리 생성(dic\_train)
  - 매핑 딕셔너리를 활용하여 train\_data, test\_data 변환
4. 최종 변환
  - 변환된 train\_data, test\_data 변환

데이터 변동성 감소로 인한 모델 성능 향상



## A. 프로젝트 소개

디스플레이 제조 공정을 위한 결함 예측 AI 모델

LG AI 해커톤

### ❖ 파생변수

#### 2. 정상/이상 데이터 분포 간 차이 기반 Feature 생성

목표: 데이터셋에서 'Normal', 'AbNormal' 간의 차이를 분석하여 새로운 피쳐 생성

사용된 기술:

- Pivot Table 활용
- MinMaxScaler를 사용한 차이값 스케일링

생성 이유:

'Normal'과 'AbNormal' 클래스 간 데이터 분포의 차이가 명확하지 않을 경우, 모델이 이상치를 정확히 분류하기 어려움

클래스 간 차이를 정량적으로 평가하고, 차이가 큰 데이터를 강조하여 모델이 학습에 중요한 정보를 더 잘 인식하도록 도움

흐름:

1. 데이터 비율 계산
  - 'Normal', 'Abnormal' 데이터 개수를 계산하고 비율(ratio) 구함
2. Pivot Table 생성
  - 각 칼럼에 대해 'Normal'과 'AbNormal'의 개수를 계산
3. 차이 계산 및 스케일링
  - 'Normal' 비율에서 'AbNormal' 비율을 빼고 절댓값 계산
  - MinMaxScaler로 차이값을 스케일링
4. 고유값 매핑 및 새로운 칼럼 생성
  - 스케일링된 차이값이 0.2 이상인 값들은 1로, 미만은 0으로 고유값의 인덱스에 매핑

클래스 간 차이를 정량화하고 중요한 정보를 강조하여 모델 성능 향상

## C. 프로젝트 소개

### 시각장애인을 위한 보조 기능 - Image Captioning

#### ❖ 연구 배경

인권위 "온라인 교육 콘텐츠 제작, 시각장애인 접근성 고려해야"



넥스트지, 시각장애인을 위한 이미지 정보 설명 프로그램 개발 출시

| 맹학교 등 일선 교육현장서 디지털 정보격차 해소 기대



정진호 기자 | ✉ 기자 페이지 구독 📖 기자의 다른기사 보기



[자세히보기] 인텔 가우디 AI 가속기 - AI 인프라의 다양한 문제 해결을 위한 대안을 제시합니다.

앞이 보이지 않는 시각장애인들에게 이미지를 파악한다는 것은 매우 어려운 일

주로 점자와 청각에 의존하여 일상생활을 하기 때문에 그로 인한 불편함 발생

Research Solution: **Image Captioning**

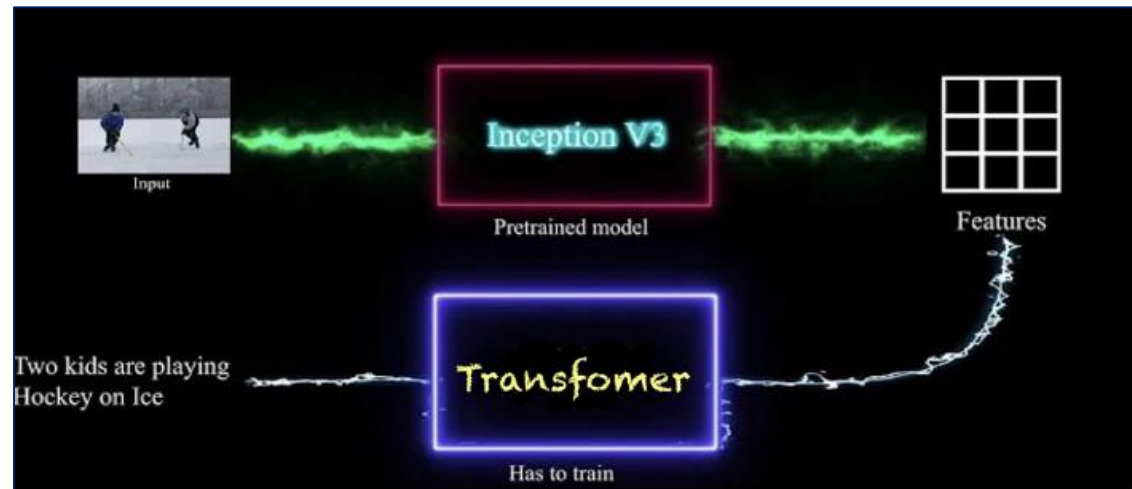
#### ❖ 연구목적

시각장애인에게 도움을 줄 수 있는 방법: Image Captioning

강의에서 배운 내용을 합쳐 응용해볼 수 있는 Multi Modal Task

여러 모델을 비교 분석하면서 성능을 정성적으로 test

#### Inception V3 + Transformer



# C. 프로젝트 소개

## 시각장애인을 위한 보조 기능 - Image Captioning

### ❖ 데이터셋



Flickr 8k Dataset

Flickr8k Dataset for image captioning.

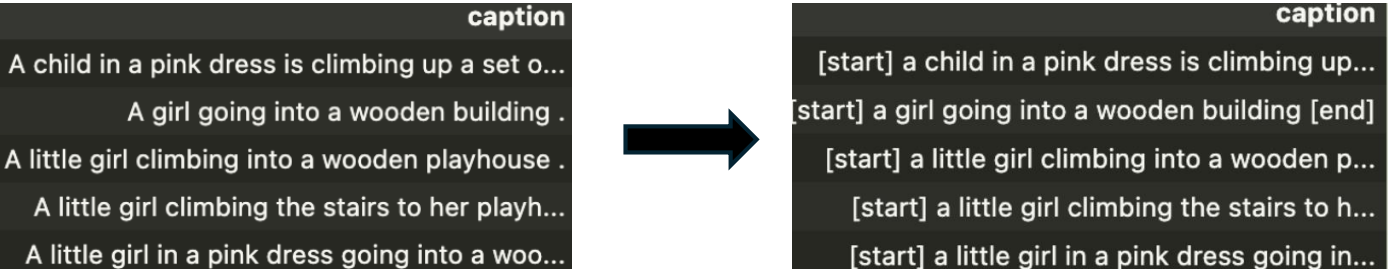


Flickr Dataset



A child in a pink dress is climbing up a set of stairs in an entry way

### ❖ Text Embedding

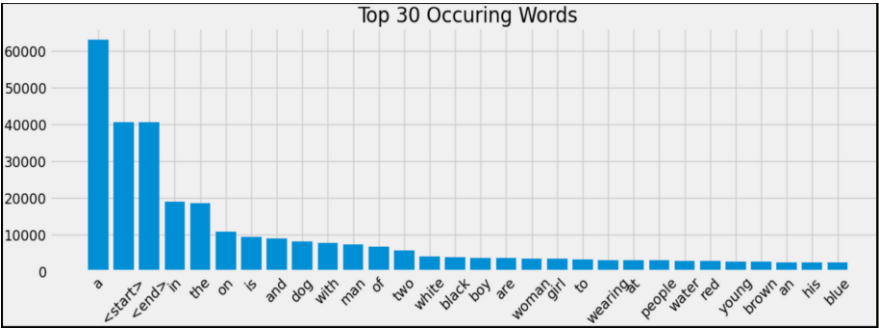


[Start], [End] 토큰 부여

<SOS : Start Of Sentence>

<EOS : End Of Sentence> 부여하여 모델이 문장을 언제 시작하고 언제 끝내야 하는지를 결정하는 역할

### ❖ Word Tokenize



단어가 최소 4번 이상 등장해야, Tokenizer 토큰화, threshold =4

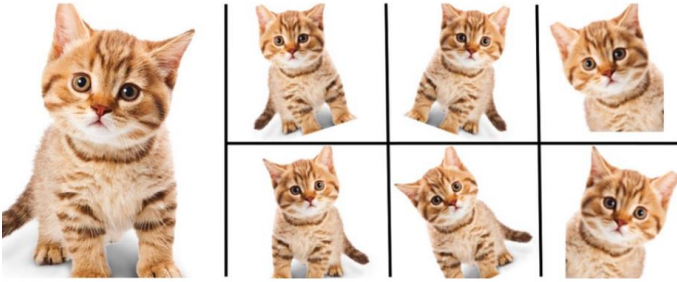
등장 횟수가 4보다 작을 시 pad(unknown) 토큰 처리

## C. 프로젝트 소개

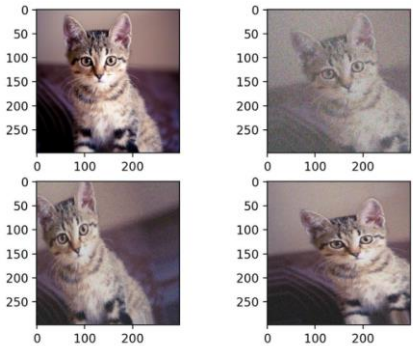
### 시각장애인을 위한 보조 기능 - Image Captioning

#### ❖ Image Augmentation & Transform

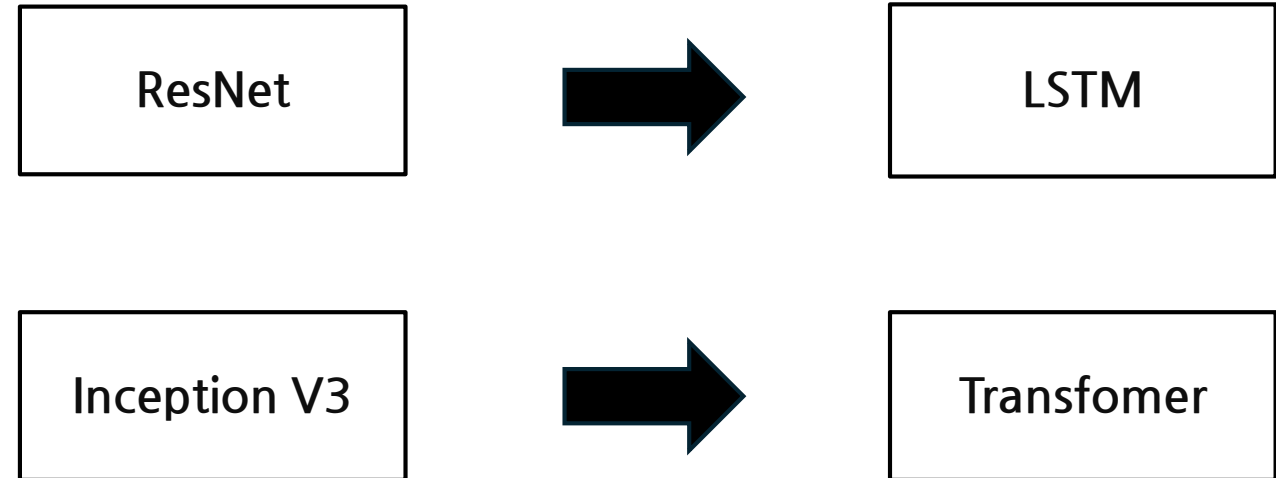
##### 1. Random Flip / Rotate



##### 2. Random Contrast



#### ❖ 구조도



-----  
여러 모델을 비교 분석 후 정성적으로 성능 테스트

# C. 프로젝트 소개

## 시각장애인을 위한 보조 기능 - Image Captioning

### ❖ 학습결과

LSTM



Two dogs are playing in the water

Transfomer

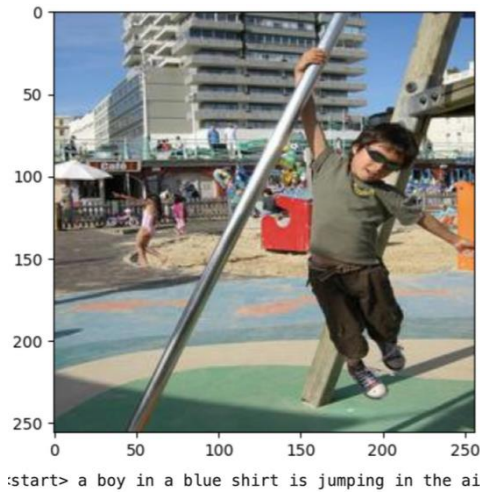


man man stands on a water

<Ground Truth>

A man is standing on an animal in the water

LSTM



a boy in a blue shirt is jumping in the air on a skateboard

Transfomer



a child swings on a pole

<Ground Truth>

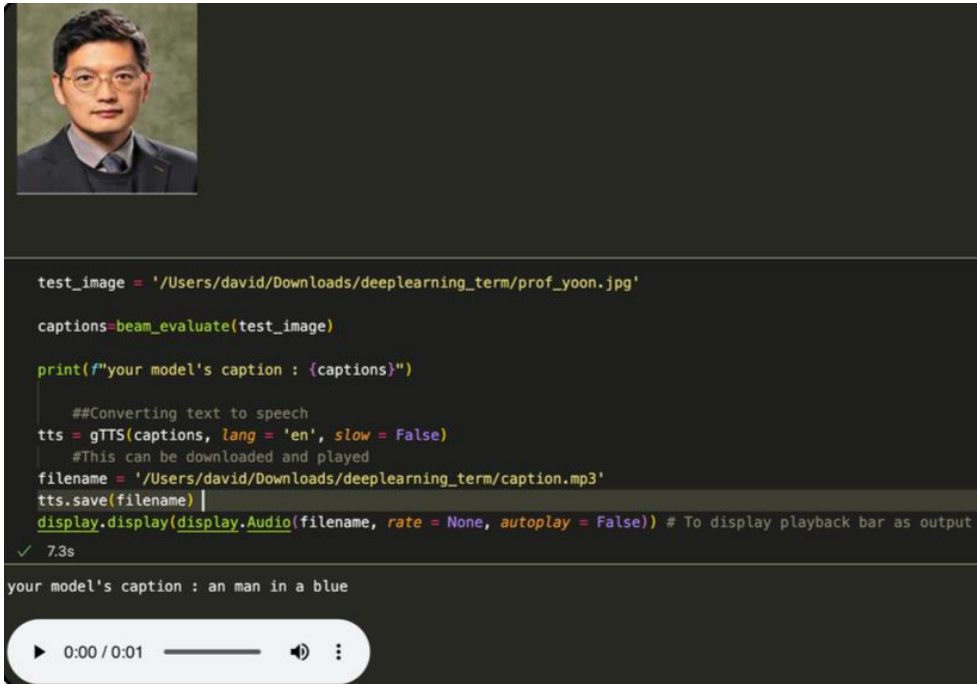
A boy hanging onto a pole



## C. 프로젝트 소개

### 시각장애인을 위한 보조 기능 - Image Captioning

#### ❖ 고도화 방향 : gTTS



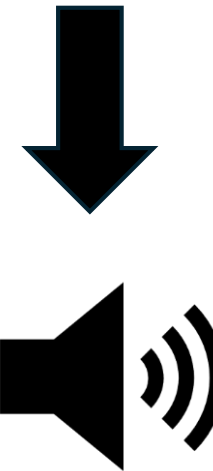
an man in a blue

<Ground Truth>

Profile of the Greatest Professor in Yonsei

caption TXT

“an man in a blue”



By using  
gTTS Library

## B. 프로젝트 소개

### 육아지원정책: KoBERTopic과 SNA를 통한 심층 분석

#### ❖ 연구 배경

##### 육아휴직 1년 6개월로 확대한다는 데... 언제? 어떻게?

법 개정 필요한 사안...국회서 법안 통과돼야  
국회 통과 시기에 따라서 내년 시행될 수도

기사입력 2023.01.12 16:51 | 최종수정 2023.01.12 16:54

우먼타임스 = 박은영 기자

최근 정부는 올해부터 육아휴직기간을 1년 6개월로 무기한 확대할 방침이다. 육아휴직기간 확대에 대한 기대감이 컸다.

그러나 막상 고용노동부 발표 내용에는...  
지, 기존 육아휴직 중인 부모도 연장 및 소  
중을 불러 일으켰다. 추가 취재를 통해 육

‘3+3 휴직’ 등 육아휴직 내년 큰 변화...올해 태어난 아이도 적용

입력 2021.09.30 (09:36) | 수정 2021.09.30 (10:10)

사회

4 9

가



- 저출산 문제를 위한 대표적인 해결책인 육아 지원 정책
- 정권 교체와 시기에 따른 육아 지원 정책들의 꾸준한 신규 수립 및 수정
- 육아 지원 정책의 대상에게 연관 정보를 더 쉽고 명확하게 전달할 수 있는 방법에 대한 의논이 필요

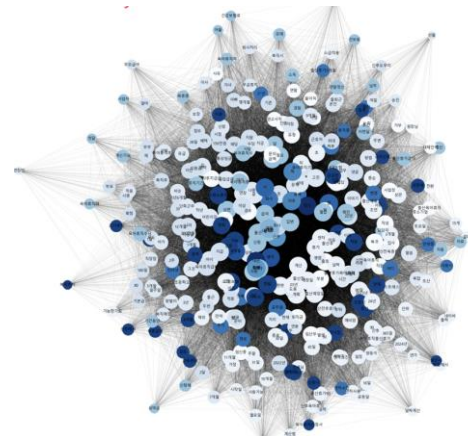
#### ❖ 연구 목표

##### Research Question

- 실제 부모들이 갖는 질문들을 바탕으로 분석을 진행하여 현실적으로 겪는 문제들을 도출
- 공통점을 가진 다양한 질문들을 연관짓고, 추가적인 질문의 재구성 진행

→ 정책의 수정 및 수립에 있어 상위포괄적인 질문을 바탕으로 효율적인 정보제공에 기여

#### SNA를 활용한 토픽 모델링 기반 질문 재구성

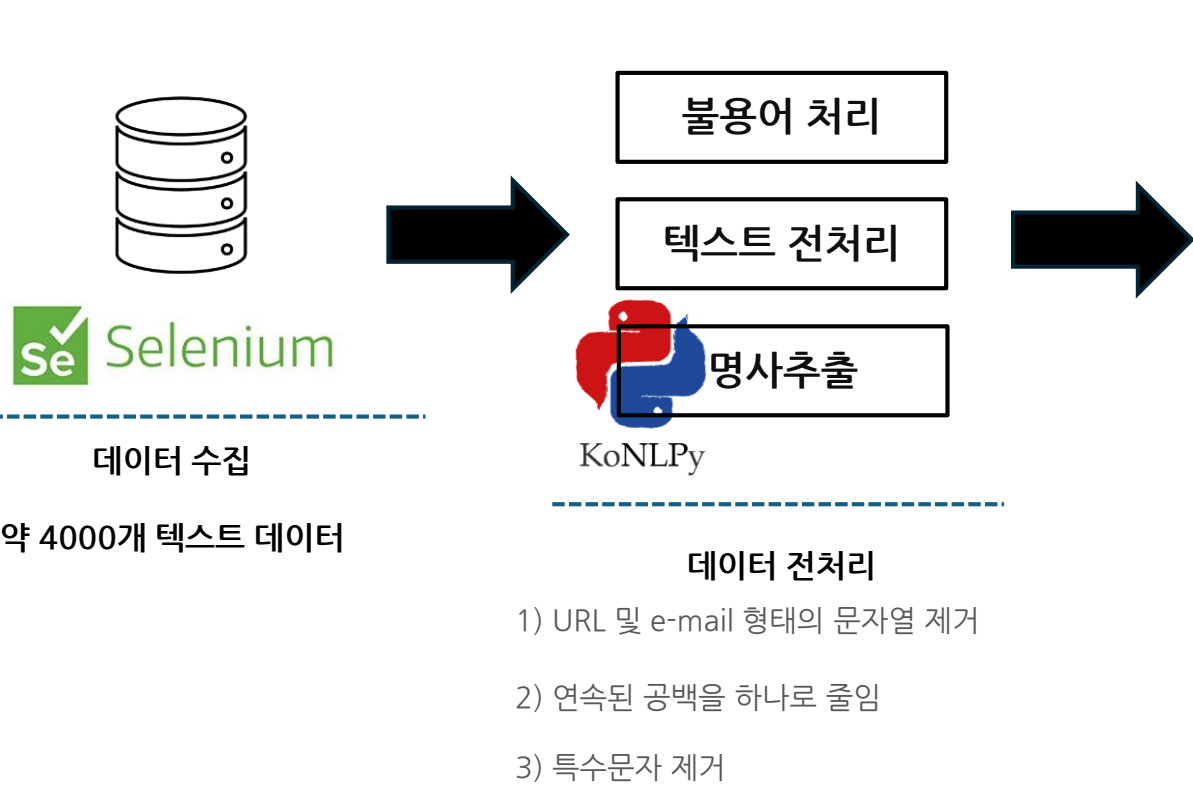


1. Cluster0 [권고사직, 계약직, 퇴직금, 사후지급금, 사업장, 대체인력]  
본인의 근로조건 상황에 따라, 충분한 육아휴직 지원을 활용하지 못할 경우, 기타 급여 처리에 관한 최적화 방안은 무엇일까?
2. Cluster1 [계획, 출산예정일, 날짜계산, 휴가, 연차, 소진]  
출산 예정일 및 연차 소진 등을 고려하여 육아휴직을 어떻게 스케줄링하는 것이 좋을까요?(출산전후휴가, 산전(후)육아휴직 등)
3. Cluster2 [어린이집, 연장반, 맞벌이, 직장맘, 신랑, 단축근무]  
맞벌이 중인 육아휴직을 고려하는 직장맘의 입장에서, 어린이집에서의 연장반 이용 유지를 위한 방법이 있을까?
4. Cluster3 [33(3+3육아휴직), 12개월, 지급, 기존, 금액, 2022년]  
3+3 부모육아휴직제의 내용과 혜택은 어떻게 진행될까?
5. Cluster6 [급여, 신청, 휴직기간, 육아휴직수당, 소득, 1년]  
육아휴직 급여 신청방법 및 신청기간에 대해 궁금해요!
6. Cluster11 [공무원, 올해, 퇴사, 변경, 인터넷, 연차수당]  
일반근로자와 달리 공무원들의 육아휴직 신청 조건 및 법 조항이 다른가요?
7. Cluster14 [고용보험, 고용노동부, 제출, 서류, 고용센터, 조기복직]  
육아휴직 신청 관련해서 필요한 서류와 급여지급은 어디에서 이뤄질까요?

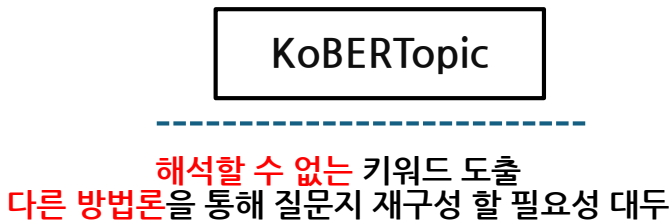
## B. 프로젝트 소개

육아지원정책: KoBERTopic과 SNA를 통한 심층 분석

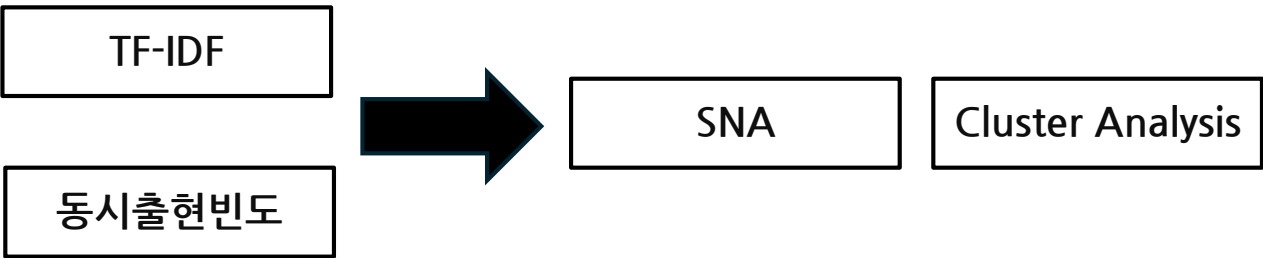
### ❖ 구조도



### 1. KoBERTopic 모델을 활용한 토픽별 키워드 기반 질문 재구성



### 2. SNA를 활용한 토픽 모델링 기반 질문 재구성

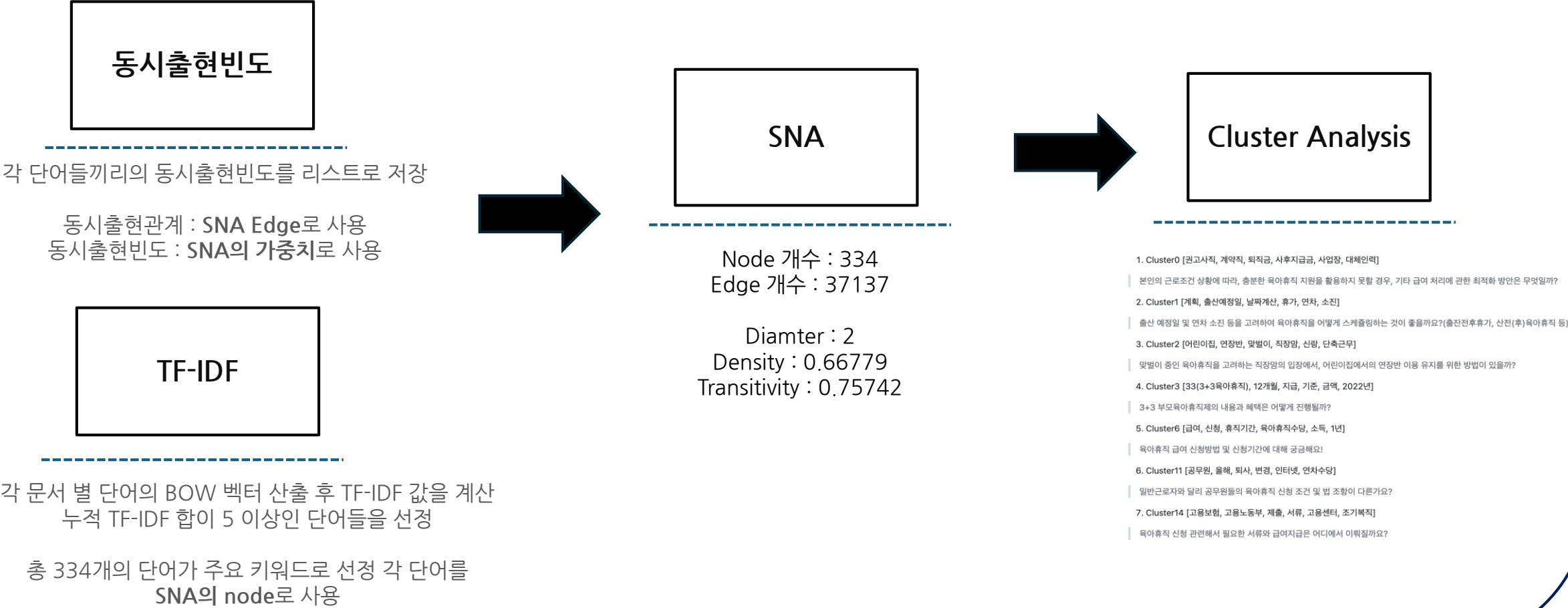


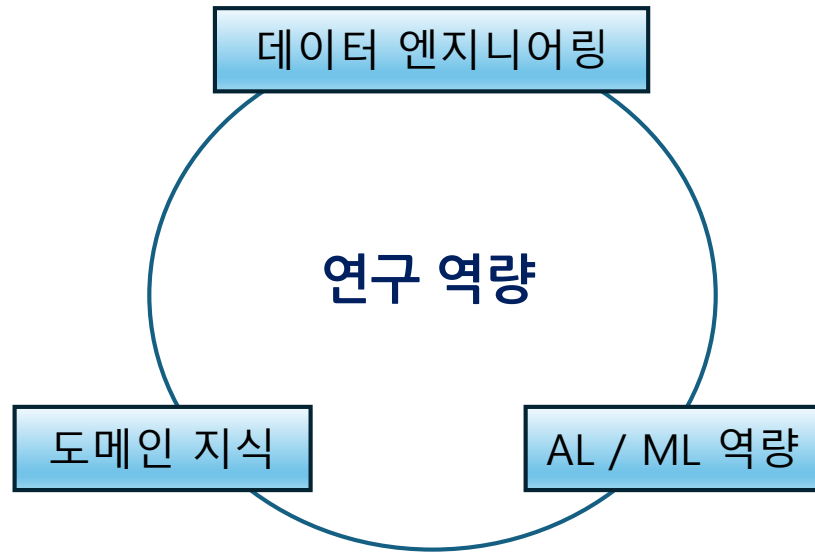


## B. 프로젝트 소개

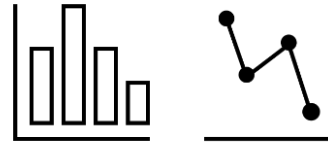
육아지원정책: KoBERTopic과 SNA를 통한 심층 분석

### 2. SNA를 활용한 토픽 모델링 기반 질문 재구성



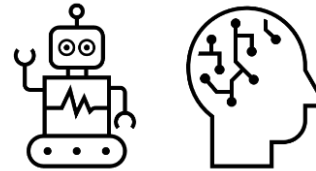


## • 데이터 엔지니어링 역량



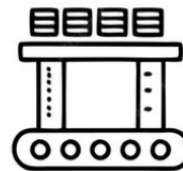
- 프로그래밍 활용
- 데이터 처리, 가공 및 해석
- 수학 / 통계적 지식

## • AI/ML 활용 역량



- 딥러닝 / 머신러닝 프레임 워크 활용(TensorFlow, Pytorch, Scikit-Learn)
- DL / ML 설계 역량

## • 도메인 지식



- 다양한 도메인 프로젝트 진행
- 실무 데이터 경험
- 물류 시스템