

2024 알테어 동지프 3기 최종 발표회

RapidMiner를 활용한 지진발생에 따른 효율적인 대비방안 탐색

상명대학교 GRIT 팀



목차

01

프로젝트 소개

- 1) 동아리 소개
- 2) 프로젝트 주제
- 3) 프로젝트 내용

04

프로젝트 결과

- 1) 결과 정리
- 2) 기대효과

02

상반기 과정 요약

- 1) 중간 발표회 내용

05

ALTAIR Tools

- 1) RapidMiner
- 2) Panopticon

03

하반기 과정 내용

- 1) 7~12월 활동 내용



01. 프로젝트 소개

1. GRIT 소개



상명대학교 GRIT

팀장

임주연

팀원

박수현, 이수진, 정성령, 한승범, 한해별

소속

상명대학교

팀 소개

'데이터를 끈기있게 파고들어

핵심과 본질을 이해하고 분석해내겠다!'

주요 활동

- 논문 탐독 및 발표
- 선정된 주제로 연구 및 논문 작성
- 알테어 동지프 2기 참여 및 수료
- 한국정보처리학회 ACK2023추계 논문 투고
- 한국정보처리학회 ASK2024 논문 투고
- KIBC 리뉴얼 프로젝트
- 알테어 동지프 3기 활동중



01. 프로젝트 소개

2. 프로젝트 주제

지진 예측은 중요하지만 다양한 변수를 고려해야 하기 때문에,
지진을 정확하게 예측하는 것은 현재 과학기술로는 어렵다는 것을 파악했습니다.

과거에 발생한 지진 데이터를 기반으로 한반도와 그 주변에서 발생한 지진의 규모들을
지역별로 파악하고 피해정도 측면에서 대피소의 위치와 개수를 파악하여
지진에 효과적으로 대비하는 방안을 탐색해보고자 합니다.

RapidMiner를 활용한 지진발생에 따른 효율적인 대비방안 탐색

01. 프로젝트 소개

3. 프로젝트 내용

Data

주소지	대피소명	위도	경도
울산광역시 북구	(구)강동중학교 운동장	35.61387	129.4441
울산광역시 북구	(구)강동초등학교 운동장	35.61588	129.4468
울산광역시 북구	(구)동해초등학교 운동장	35.5818	129.4547
부산광역시 강서구	(구)외항포가는길	35.02849	128.8179
경상북도 포항시	1박2일펜션 앞	36.23439	129.3811
경상북도 울진군	7번국도 옆 공장부지	36.67498	129.4403
경상북도 영덕군	7번국도변	36.62499	129.4113
경상북도 영덕군	7번국도변	36.57654	129.3997
부산광역시 사하구	BMC아파트 테니스장	35.06456	128.9789
경상북도 포항시	MGM 그랜드호텔 앞	36.00014	129.5685
강원도 동해시	가세마을야산	37.52753	129.122
강원도 삼척시	갈남마을회관뒤공터	37.27782	129.3263
강원도 삼척시	갈남전망대	37.28108	129.3229
경상북도 영덕군	갈매기횃집뒤	36.40888	129.4288

- 국내 지진데이터
- 실내 구호소 및 옥외대피소
- 전국 지진해일 대피소
- 재난문자 전송 데이터

기존 지진 발생 사례의 피해 관련 데이터를 활용

변수별 지진의 피해정도를 다양한 모델을 이용해 예측

Panopticon으로 지역별 시각화 진행

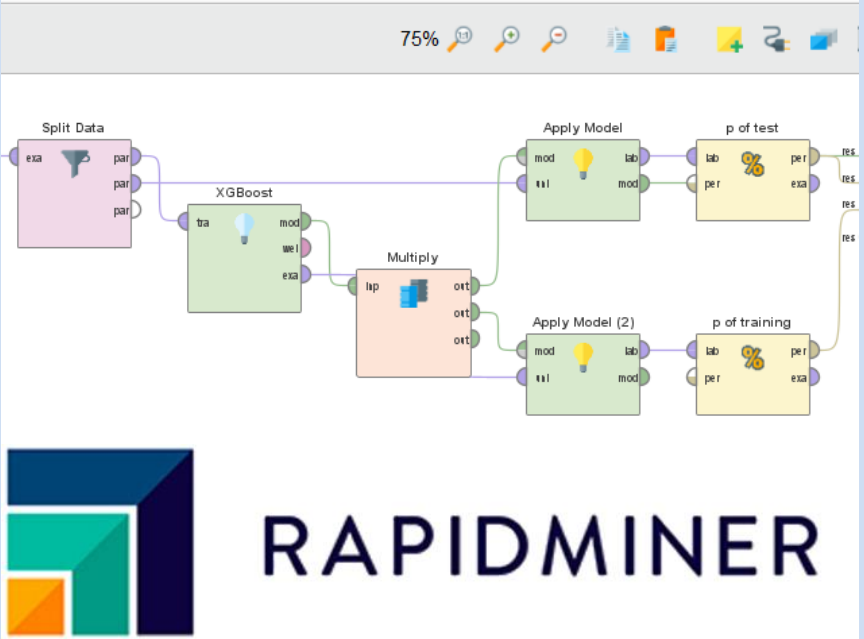
이를 통해 지진발생에 따른 효율적인 대비방안을 탐색하고자 합니다.


이에 따라, 피해정도가 심한 지역과 실내구호소, 대피소의 위치, 개수를 파악하여 어떤 점 등을 개선해야 할 지 결정할 수 있습니다.

01. 프로젝트 소개

3. 프로젝트 내용

RapidMiner





변수에 따른 지진
피해정도를 RapidMiner의
다양한 모델로 예측 진행

기존 지진 발생 사례의 피해 관련 데이터를 활용

변수별 지진의 피해정도를 다양한 모델을 이용해 예측

Panopticon으로 지역별 시각화 진행

이를 통해 지진발생에 따른 효율적인 대비방안을 탐색하고자 합니다.

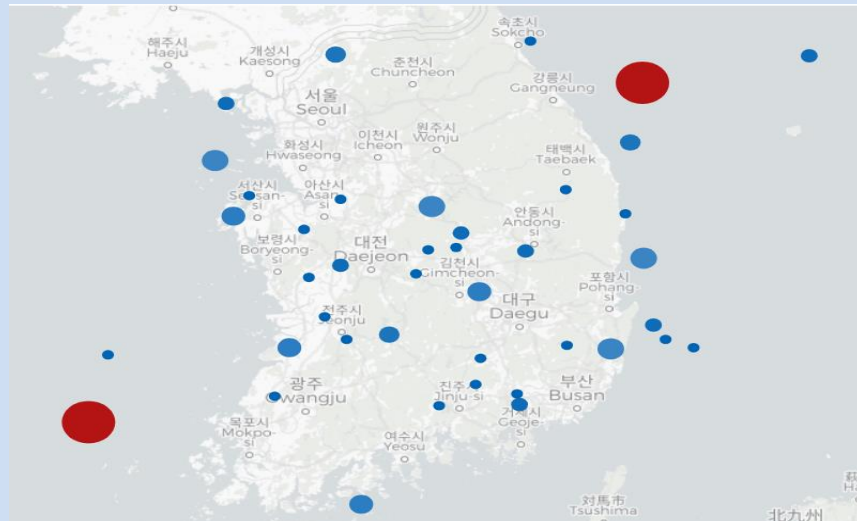
이에 따라, 피해정도가 심한 지역과 실내구호소, 대피소의 위치, 개수를 파악하여 어떤 점 등을 개선해야 할 지 결정할 수 있습니다.



01. 프로젝트 소개

3. 프로젝트 내용

Panopticon



Altair Panopticon™

Panopticon을 활용한
결과 시각화

기존 지진 발생 사례의 피해 관련 데이터를 활용
변수별 지진의 피해정도를 다양한 모델을 이용해 예측

Panopticon으로 지역별 시각화 진행

이를 통해 지진발생에 따른 효율적인 대비방안을 탐색하고자 합니다.
이에 따라, 피해정도가 심한 지역과 실내구호소, 대피소의 위치, 개수를 파악하여 어떤
부분을 개선해야 할 지 결정할 수 있습니다.



01. 프로젝트 소개

3. 프로젝트 내용

Data

- 국내 지진데이터
- 실내 구호소 및 옥외대피소
- 전국 지진해일 대피소
- 재난문자 전송 데이터

RapidMiner

변수에 따른 지진
피해정도를 RapidMiner의
다양한 모델로 예측 진행

Panopticon

Panopticon을 활용한
결과 시각화

이를 통해 지진발생에 따른 효율적인 대비방안을 탐색하고자 합니다.

이에 따라, 피해정도가 심한 지역과 실내구호소, 대피소의 위치, 개수를 파악하여 어떤 부분을 개선해야 할 지 결정할 수 있습니다.

01. 프로젝트 소개

3. 프로젝트 내용

1) 효율적인 지진 대비 방안 마련

피해정도를 예측하는 모델을 만들고 지역별 시각화 통해, 각 지역별 대피소 개수와 위치를 비교하고 피해정도를 기준으로 효율적인 대비 방안을 제안해 보았습니다.

2) 데이터 활용 역량 강화

RapidMiner와 Panopticon으로 다양한 모델을 적용하고 시각화 하여 비교하는 과정을 통해 데이터 분석 역량을 강화할 수 있었습니다.



02. 상반기 과정 요약

• 3월

- RapidMiner의 Data Science Professional with RapidMiner 수강
- RapidMiner 프로그램 설치
- RapidMiner 커뮤니티 활용

• 4월

- 프로젝트에 사용할 지진 관련 데이터 수집
- 국내 지진 데이터
 - 전국 실내구호소/옥외대피소 데이터
 - 전국 지진해일 대피소 데이터
 - 재난문자 전송 데이터

• 5월

- RapidMiner - 데이터 전처리 작업
- 이상치 확인
 - Min-Max 정규화
 - 종속변수 설정: '피해정도'

• 6월

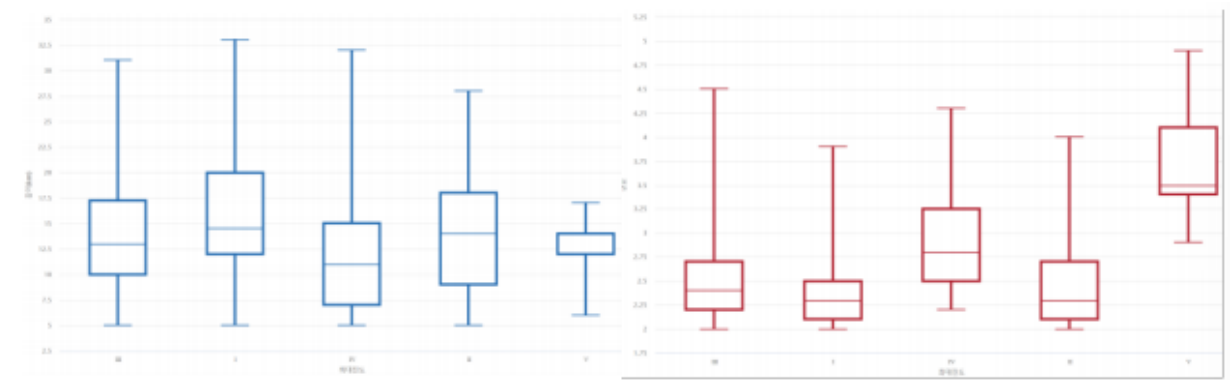
- 데이터를 훈련/테스트 set으로 분할
- 훈련 set 학습을 위해, 다중선형회귀식 도출

• RapidMiner의 Data Science Professional with RapidMiner 수강



• 지진 관련 데이터 수집 및 전처리

- * 피해정도 예측을 위한 데이터
 - 국내 지진 데이터
 - 재난문자 전송 데이터
- * 지진 대피소 현황 데이터
 - 전국 실내구호소 데이터
 - 전국 옥외대피소 데이터
 - 전국 지진해일 대피소 데이터



- * 이상치 확인 - Box Plot
- * Min-Max 정규화 - 단위 통일
- * 종속변수 설정 : $\text{피해 정도} = \text{SQRT}((\text{규모}^2 + \text{진도}^2) / 2)$
- '피해정도'의 기준: 규모와 진도를 활용 \rightarrow n차원 좌표평면에서 L2 NORM 값을 기준으로 산출

02. 상반기 과정 요약

• 3월

- RapidMiner의 Data Science Professional with RapidMiner 수강
- RapidMiner 프로그램 설치
- RapidMiner 커뮤니티 활용

• 4월

- 프로젝트에 사용할 지진 관련 데이터 수집
- 국내 지진 데이터
 - 전국 실내구호소/옥외대피소 데이터
 - 전국 지진해일 대피소 데이터
 - 재난문자 전송 데이터

• 5월

- RapidMiner - 데이터 전처리 작업
- 이상치 확인
 - Min-Max 정규화
 - 종속변수 설정: '피해정도'

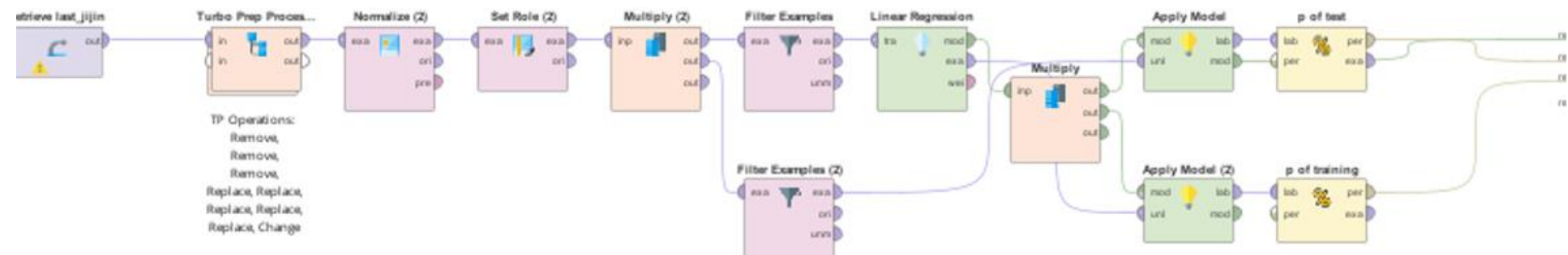
• 6월

- 데이터를 훈련/테스트 set으로 분할
- 훈련 set 학습을 위해, 다중선형회귀식 도출

- 수집한 데이터를 Training set과 Test set으로 분할

- Training set 학습을 위한 다중선형회귀식 도출

- 회귀식: $1.722 * \text{규모} - 0.240 * \text{깊이(km)} + 2.610 * \text{최대진도} + 1.841$



Result History PerformanceVector (P of training)

LinearRegression

1.722 * 규모
 - 0.240 * 깊이(km)
 + 2.610 * 최대진도
 + 1.841

Result History PerformanceV

PerformanceVector

PerformanceVector:
 root_mean_squared_error: 0.270 +/- 0.000
 squared_correlation: 0.803

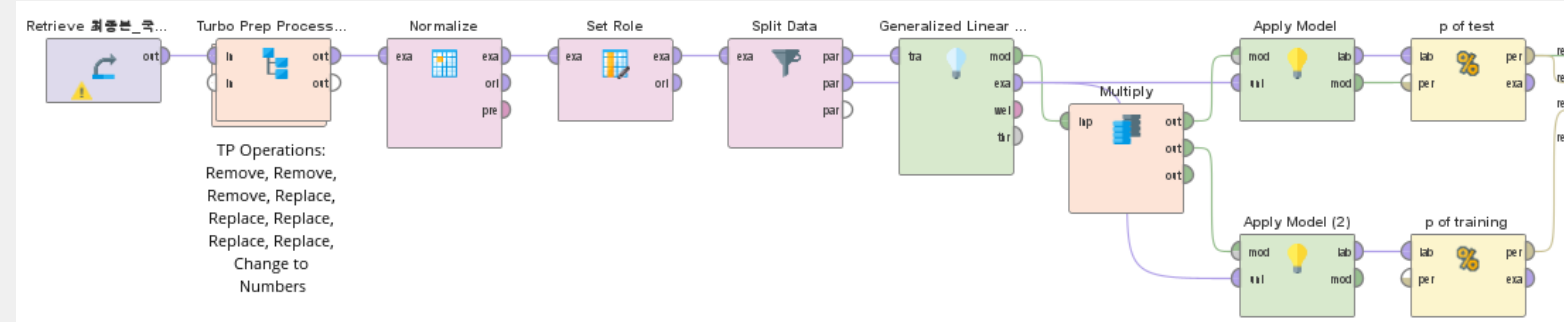
03. 하반기 과정 내용

7~8월 활동

- RapidMiner를 활용한 다양한 모델링

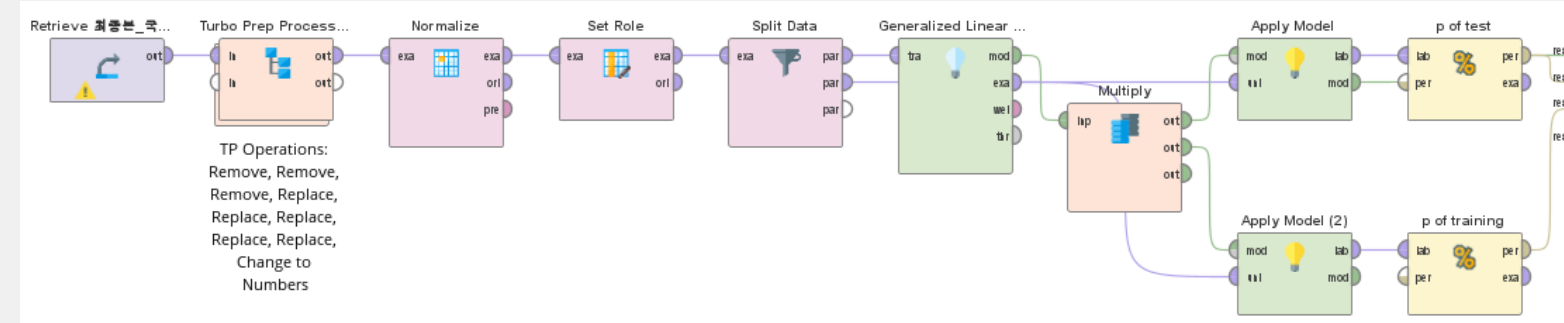
Ridge Regression

Operator: GLM
 Parameter: $\alpha = 0$
 Accuracy: 0.811



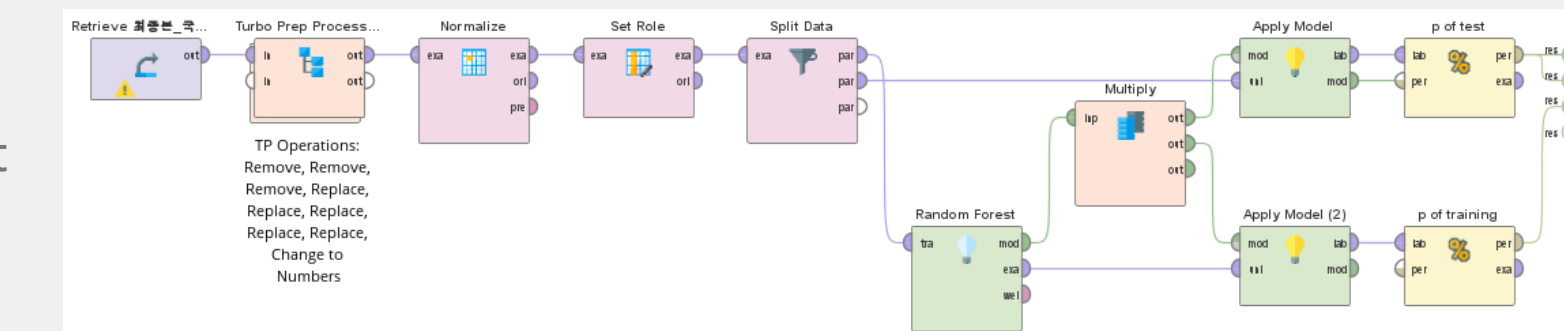
Lasso Regression

Operator: GLM
 Parameter: $\alpha = 1$
 Accuracy: 0.811



Random Forest

Operator: Random Forest
 Accuracy: 0.985



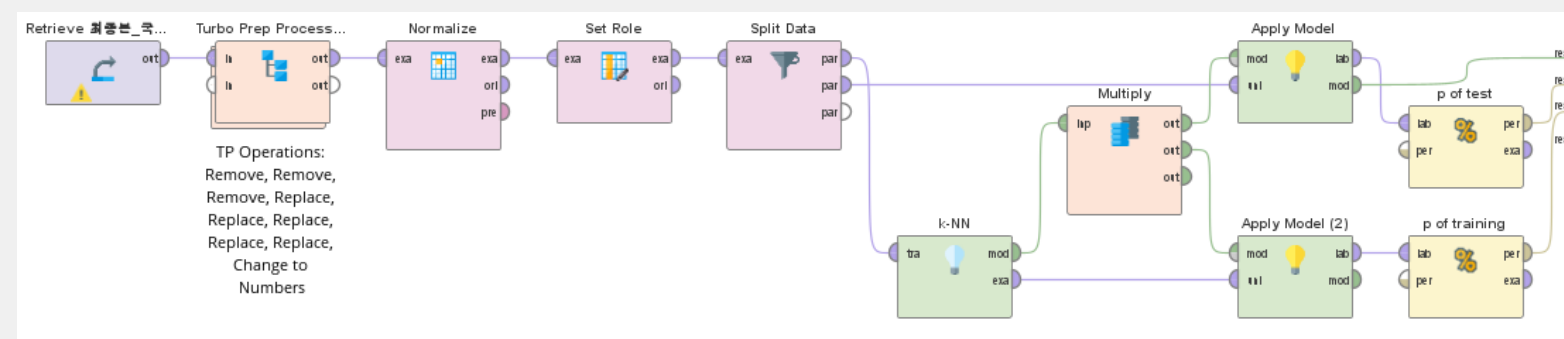
03. 하반기 과정 내용

7~8월 활동

- RapidMiner를 활용한 다양한 모델링

• KNN

Operator: K-NN
Accuracy: 0.966



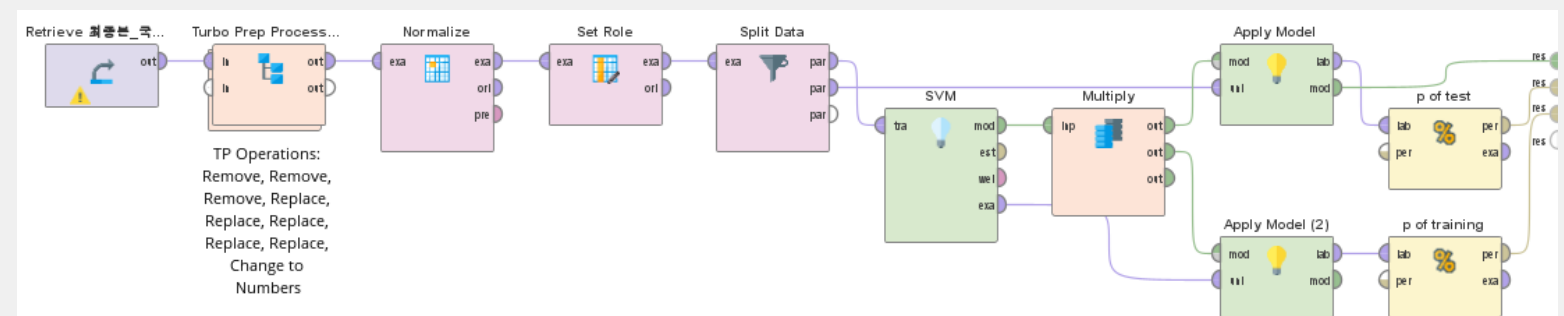
• SVM

Operator: SVM

Kernel type: dot
Accuracy: 0.8

Kernel type: ANOVA
Accuracy: 0.962

Kernel type: radial
Accuracy: 0.921



03. 하반기 과정 내용

7~8월 활동

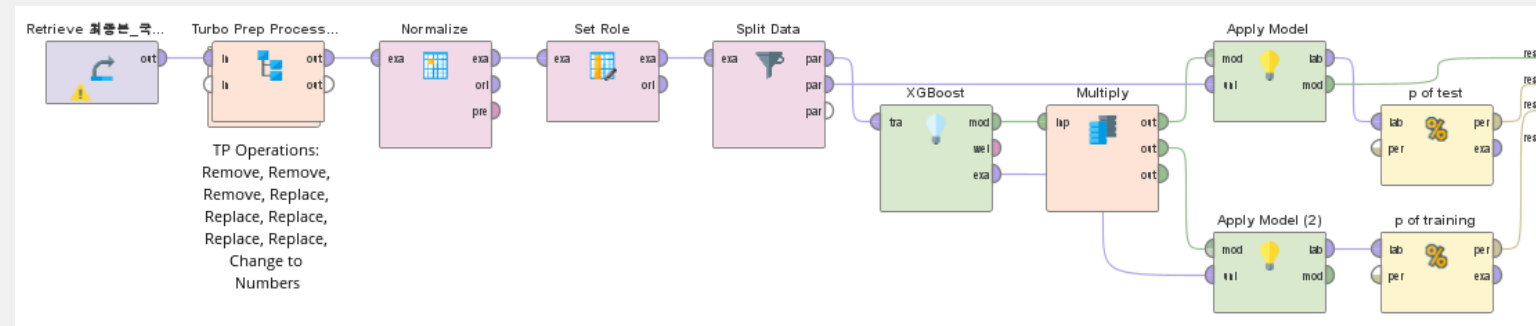
- RapidMiner를 활용한 다양한 모델링

- XGBoost

Operator: XGBoost

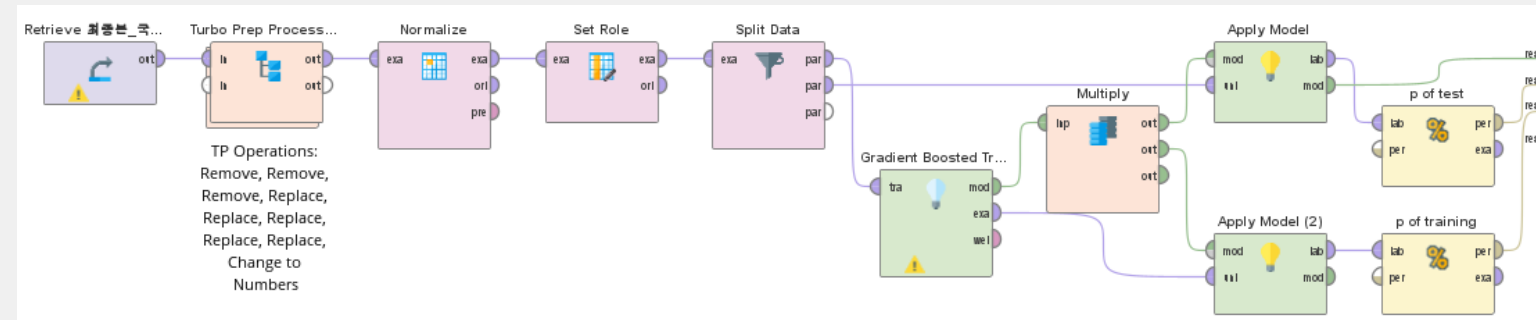
Booster: tree
Accuracy: 1.0

Booster: linear
Accuracy: 0.8



- Gradient Boost

Operator: Trees
Accuracy: 0.978



03. 하반기 과정 내용

7~8월 활동

- RapidMiner를 활용한 다양한 모델링

모델	정확도	학습 파라미터
Ridge Regression	0.811	a=0
Lasso Regression	0.811	a=1
RandomForest	0.985	
K-NN	0.966	
SVM	0.8	dot
	0.962	ANOVA
	0.921	radial
XGBoost	1.0	tree
	0.8	Linear



03. 하반기 과정 내용

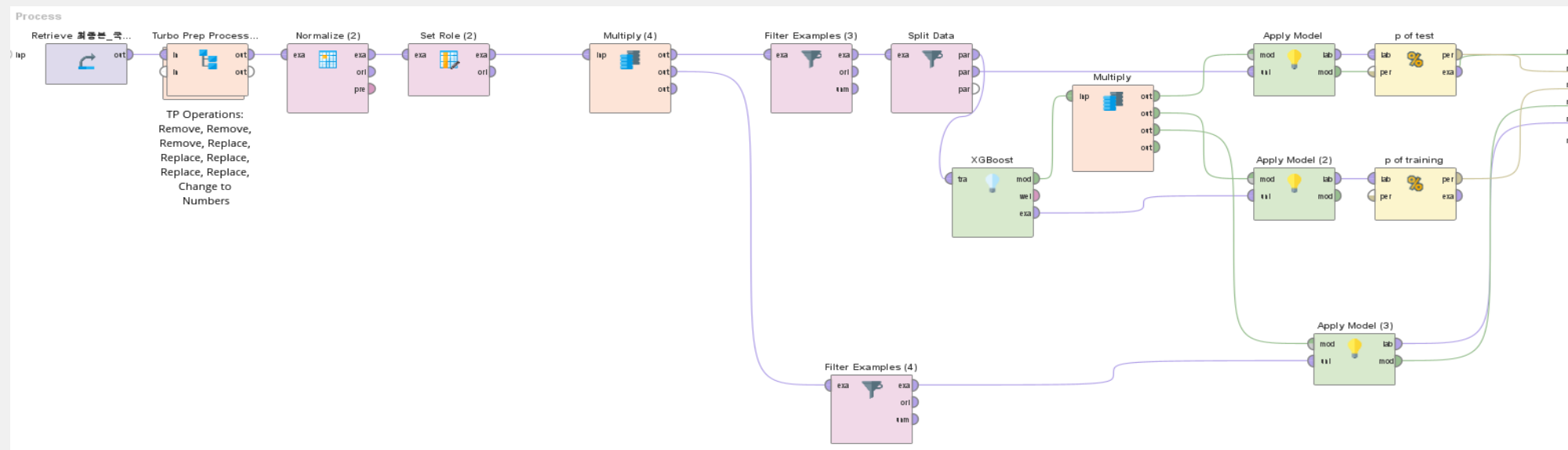
9월 활동

- 지진 피해정도 예측 모델

• XGBoost

Operator: XGBoost
 Booster: tree
 Filter

- Training set 분할 기준: '피해정도' is not missing (2019 ~ 2023)
- Test set 분할 기준: '피해정도' is missing (2024)



실제 값	예측 값
1.838	1.838
1.838	1.838
1.838	1.838
1.645	1.645
1.581	1.581
3.036	3.054
2.000	2.005
1.773	1.775
1.709	1.707
2.102	2.136



03. 하반기 과정 내용

10월 활동

- ALTAIR Panopticon 설치 및 로그인

- ALTAIR Panopticon

방대한 양의 데이터를 기반으로 통찰력 있고, 필요한 관점을 제공하는 시각적 분석 대시보드 툴

- Panopticon 설치 방법

메뉴얼) ALTAIR 온라인 스토어 - 자료실

Altair Panopticon Installation File

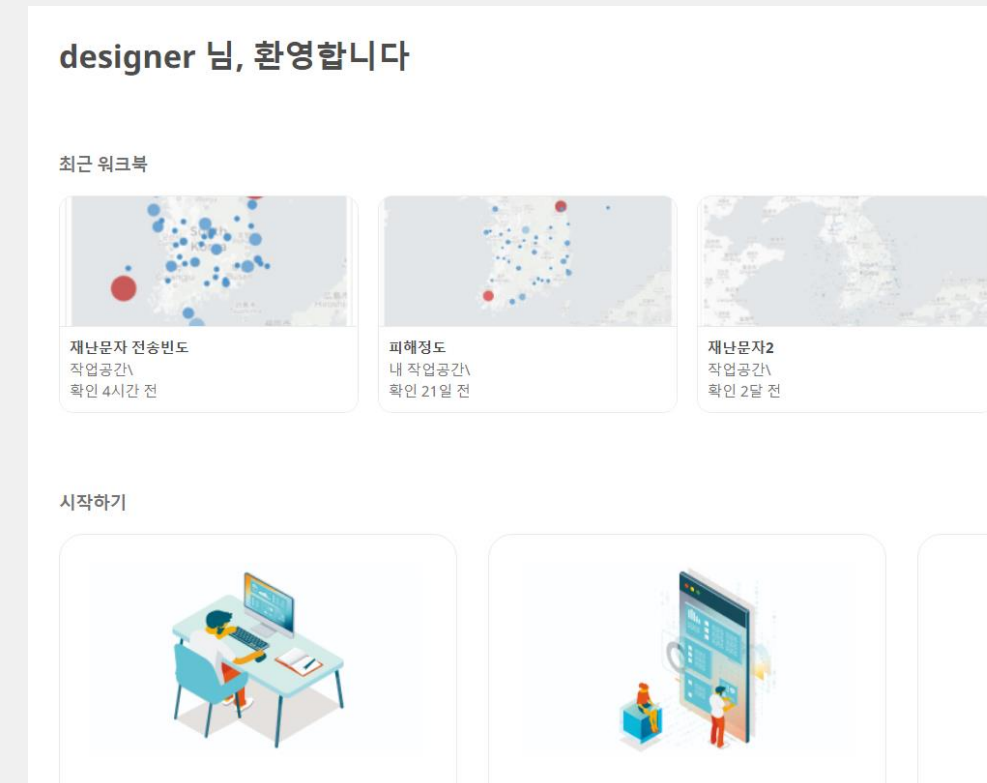
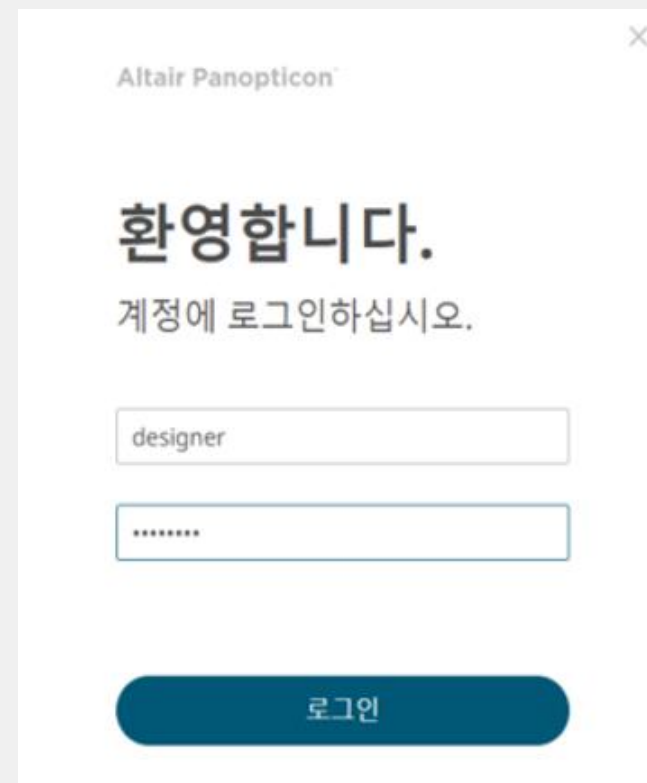
License Manager Download

Altair License Utility 설정

start.bat 파일 클릭

localhost:8080/panopticon/ 접속

designer로 로그인

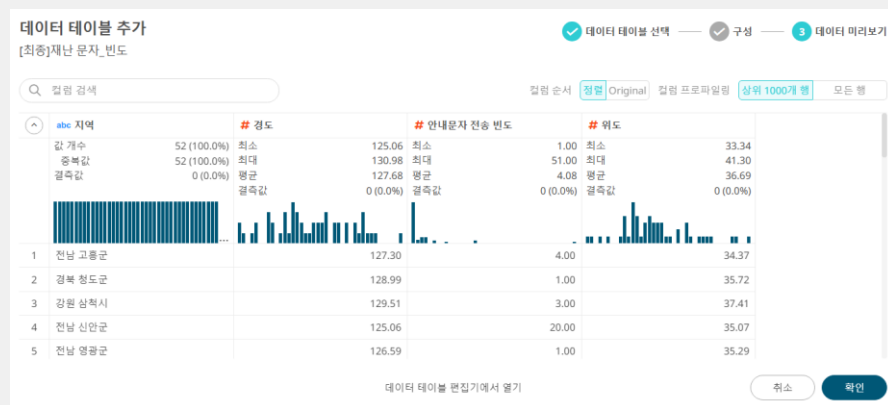


03. 하반기 과정 내용

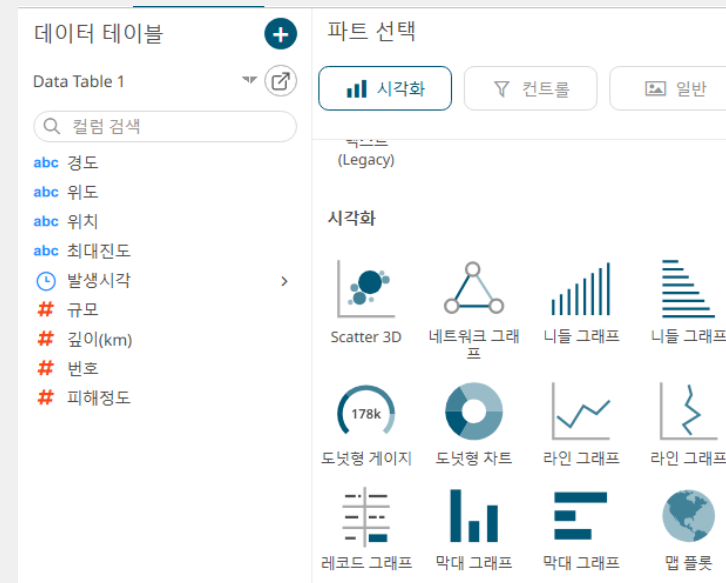
10월 활동

- ALTAIR Panopticon - Dashboard

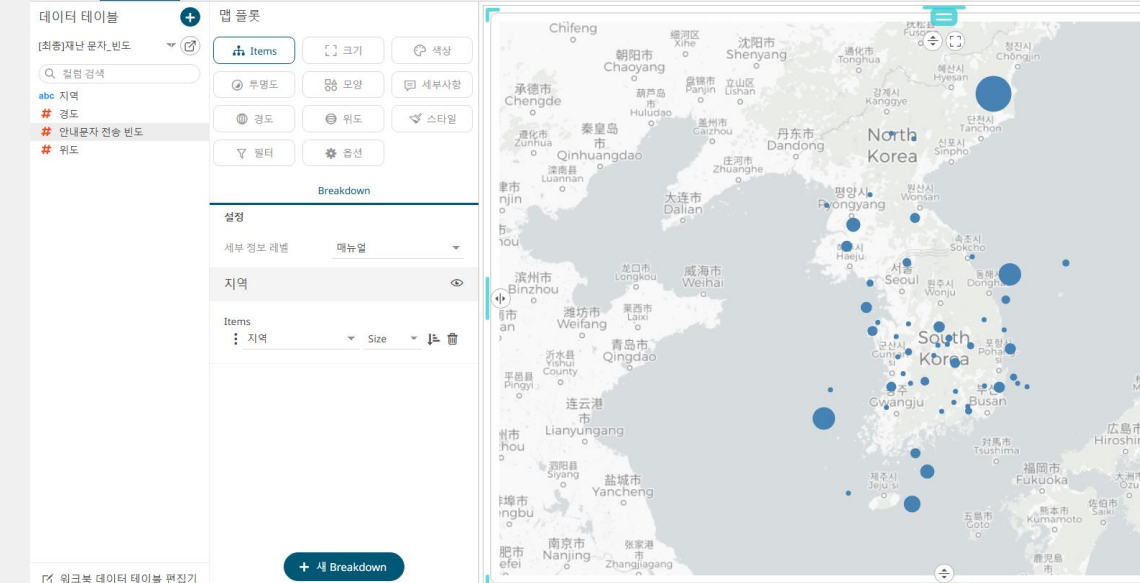
- Workbook 및 Dashboard 생성
- 데이터 테이블 추가



- 파트 선택 - 시각화



- 맵 플롯 - 경도, 위도, Items 등 옵션 설정



데이터 테이블로 사용할 데이터를 업로드 → 파트 선택에서 시각화 방식을 선택 → 지도 시각화를 위해 맵 플롯 선택 → 위도, 경도를 맵 플롯 설정에 drag and drop → Items에는 지역 → 크기와 색상은 나타내고 싶은 변수를 drag and drop

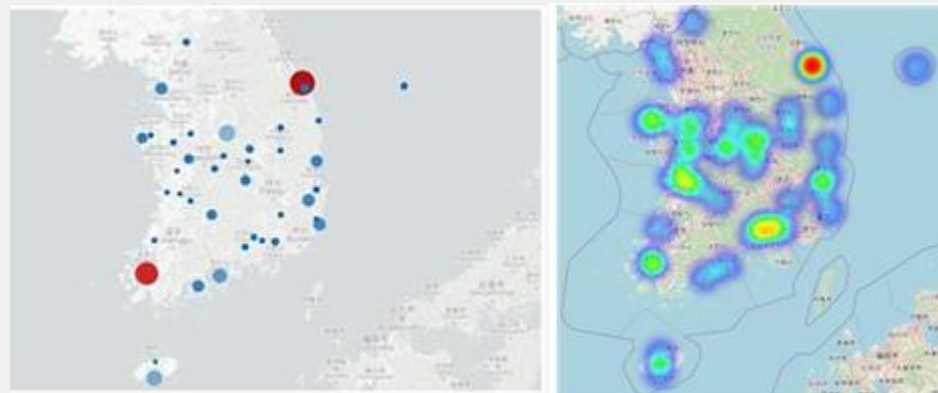
03. 하반기 과정 내용

11월 활동

- Panopticon과 Python을 이용한 시각화

- 국내 지진 피해 정도 합계 시각화

- 1) 수집 나라: 한국
- 2) 수집 기간: 2019~2024 → 2022~2024
- 3) 위치 범위: 구체적 위/경도 → 시군구 단위 위/경도
- 4) 시군구를 기준으로 합계로 집계



시군구	#	피해정도	city	#	longitude	#	latitude
동해시	44.03360244	동해시	129.1166333	37.52193056			
신안군	37.50348229	신안군	126.3817306	34.78981111			
괴산군	19.69661078	괴산군	127.7888306	36.81243056			
서귀포시	17.75071209	서귀포시	126.5125556	33.25235			
여수시	14.39906218	여수시	127.6643861	34.75731111			
경주시	11.33135444	경주시	129.2270222	35.85316944			
고흥군	10.4001556	고흥군	127.2870556	34.60806944			
웅진군	10.28983913	웅진군	126.6388889	37.443725			
영덕군	10.17449207	영덕군	129.3683556	36.41210278			
태안군	8.394375048	태안군	126.299975	36.74266667			
장수군	7.851203031	장수군	127.5233	35.64429722			
김천시	7.793393563	김천시	128.1158	36.13689722			
공주시	6.992127207	공주시	127.1211194	36.44361389			
삼척시	6.438698316	삼척시	129.1674889	37.44708611			
강원	6.27148067						
문경시	4.956324243	문경시	128.1890194	36.58363056			
창원시	4.673013281	창원시	128.6401544	35.2540033			
연천군	4.15298927	연천군	127.0770667	38.09336389			



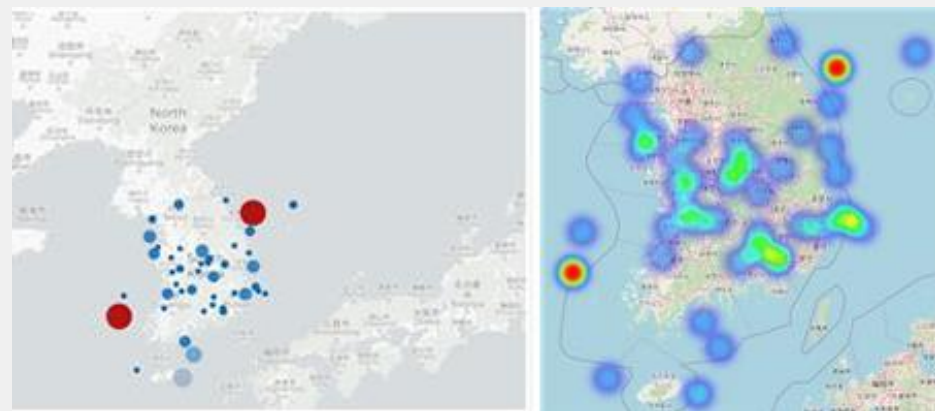
03. 하반기 과정 내용

11월 활동

- Panopticon과 Python을 이용한 시각화

- 재난문자 전송 빈도 시각화

- 1) 수집 나라: 전세계 → 한국
- 2) 수집 기간: 2006~2024 → 2022~2024
- 3) 위치 범위: 구체적 위/경도 → 시군구 단위 위/경도
- 4) 시군구를 기준으로 빈도로 집계



	A	B	C	D
1	지역	안내문자 전송 빈도	위도	경도
2	전남 신안군	20	35.07	125.06
3	강원 동해시	20	37.9	129.61
4	제주 서귀포시	11	33.34	127.22
5	전남 여수시	8	34	127.59
6	경북 경주시	5	35.69	129.35
7	경북 영덕군	5	36.45	129.62
8	인천 옹진군	5	37.26	126.1
9	충북 괴산군	5	36.88	127.88
10	전남 고흥군	4	34.37	127.3
11	충남 태안군	4	36.8	126.25
12	경북 김천시	4	36.17	128.27
13	전북 부안군	4	35.7	126.71
14	강원 삼척시	3	37.41	129.51
15	경기 연천군	3	38.13	127.09
16	전북 장수군	3	35.81	127.53
17	충남 공주시	2	36.39	127.13
18	경남 창원시	2	35.22	128.6
19	경북 울릉군	2	38.12	130.98
20	경북 포항시	2	35.89	129.7
21	경북 안동시	2	36.51	128.65
22	인천 강화군	2	37.73	126.19
23	경북 문경시	2	36.66	128.12
24	경북 청도군	1	35.72	128.99
25	전남 영광군	1	35.29	126.59



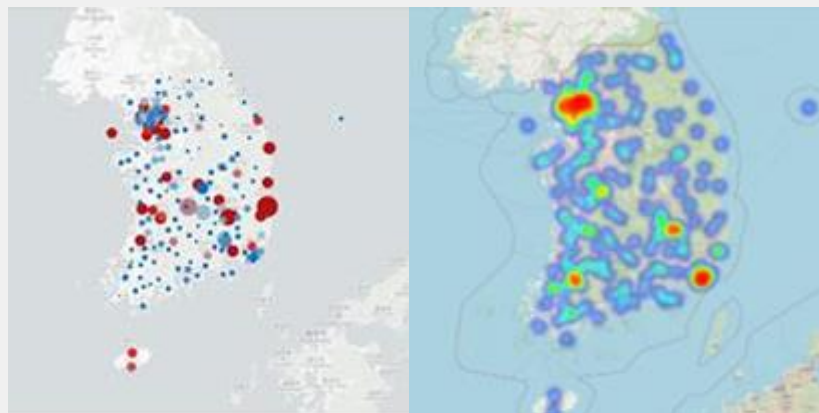
03. 하반기 과정 내용

11월 활동

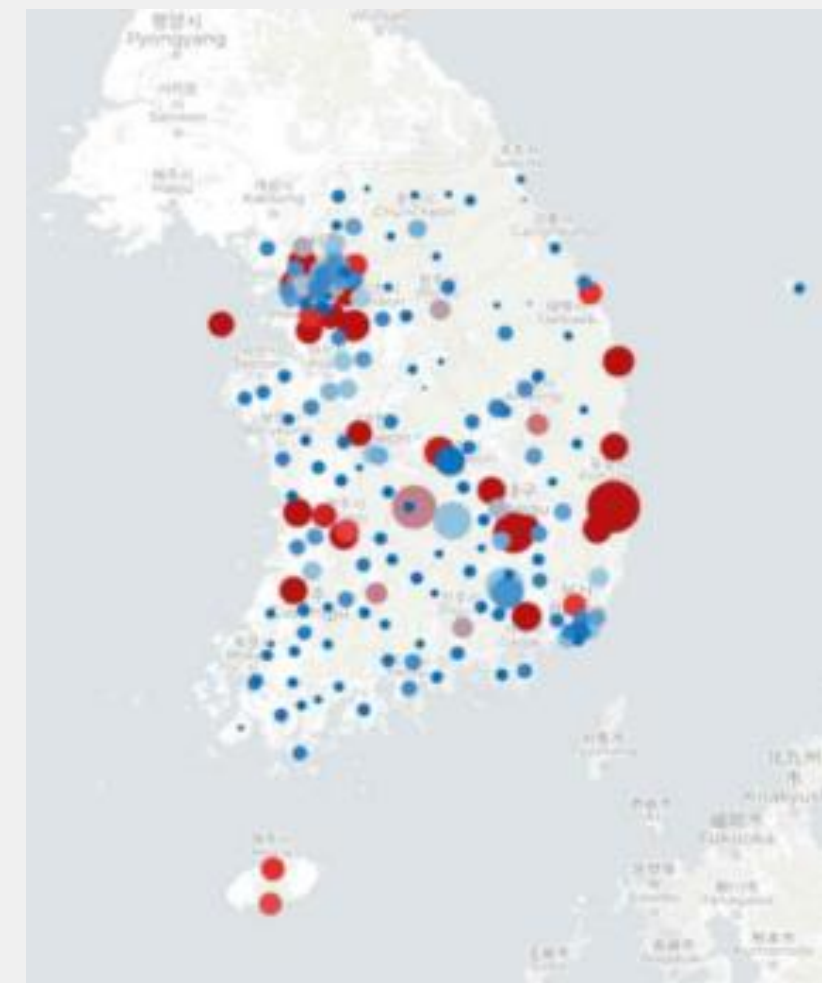
- Panopticon과 Python을 이용한 시각화

- 대피소 개수 시각화

- 1) 데이터: 옥외대피소 + 지진겸용임시주거시설 + 지진해일대피소(실내구호소)
- 2) 수집 나라: 한국
- 3) 위치 범위: 구체적 위/경도 → 시군구 단위 위/경도
- 4) 시군구를 기준으로 개수로 집계



주소지	시군구	# 개수	# 위도	# 경도
경상	포항시	501	36.00568611	129.3616667
대구시	달서구	306	35.82692778	128.5350639
경기	용인시	188	37.23147778	127.2038444
경상	울진군	176	36.99018611	129.4027861
인천시	부평구	174	37.50784204	126.7219068
경상	창원시	156	35.2540033	128.6401544
경상	경주시	154	35.85316944	129.2270222
서울	강남구	153	37.514575	127.0495556
전라	군산시	152	35.96464167	126.7388444
전라	전주시	151	35.80918889	127.1219194
대구시	수성구	144	35.85520833	128.6328667
전라	고창군	142	35.43273889	126.7041083
경상	영덕군	141	36.41210278	129.3683556
경기	고양시	140	37.65590833	126.7770556
경기	수원시	140	37.30101111	127.0122222
경상	구미시	140	36.11655	128.3467778
서울	강서구	136	37.54815556	126.851675
경기	화성시	129	37.19681667	126.8335306



03. 하반기 과정 내용

12월 활동

- Panopticon 시각화를 통한 결론 도출

- 종속변수 '피해정도'의 타당성 확인

'지진 피해정도': 1, 2, 4위 → '지진 재난 안내 문자 빈도': 2, 1, 3위

⇒ '지진 피해정도'와 '지진 재난 안내 문자 빈도'의 분포가 매우 유사하게 나타남을 통해 종속변수의 타당성 확인



<지진 피해 정도>



<재난문자 전송 빈도>



03. 하반기 과정 내용

12월 활동

• Panopticon 시각화를 통한 결론 도출

• 지진 피해정도에 따른 대피소 위치/개수 적합성 판단

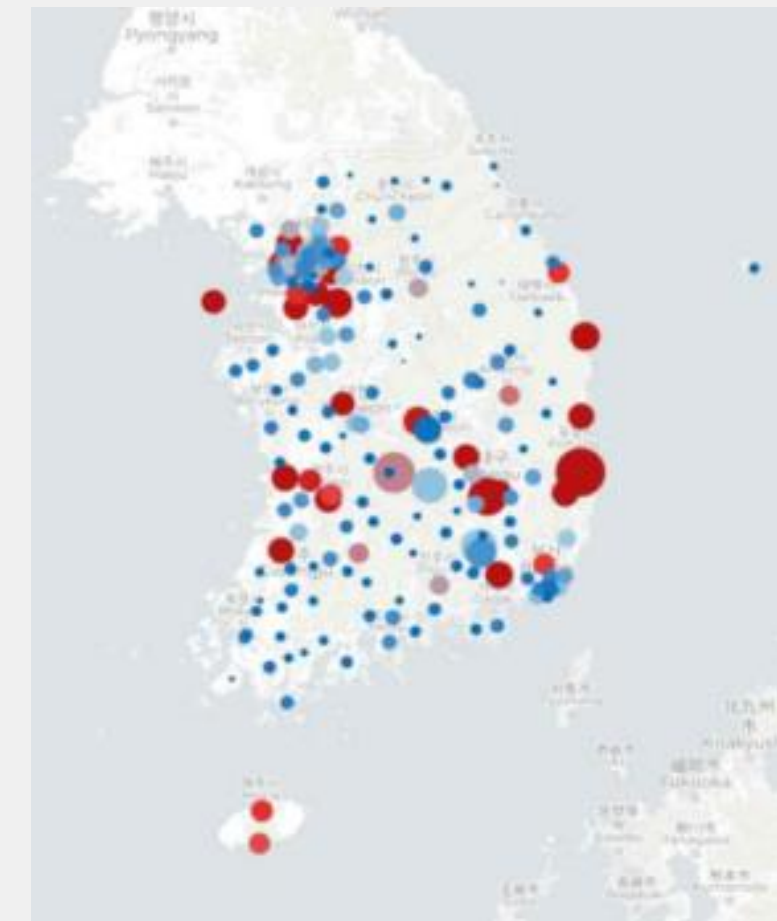
'지진 피해정도': 1, 2, 3위 → '대피소 개수': 122, 127, 221위

- 높은 '지진 피해정도' 를 갖는 지역의 대피소 개수가 중하위권에 존재함
- 서울, 수도권, 부산, 광주 등의 지역은 피해정도가 낮고 대피소 개수가 많음

⇒ 지역별 지진 피해정도에 따른 시각화와 재난문자 전송 빈도 시각화의 붉은 색으로 표시된 지역인 동해시와 신안군의 지진 대피소 최적 입지 선정 및 마련과 대처방안을 더욱 강구해야 함



<지진 피해 정도>



<대피소 개수>



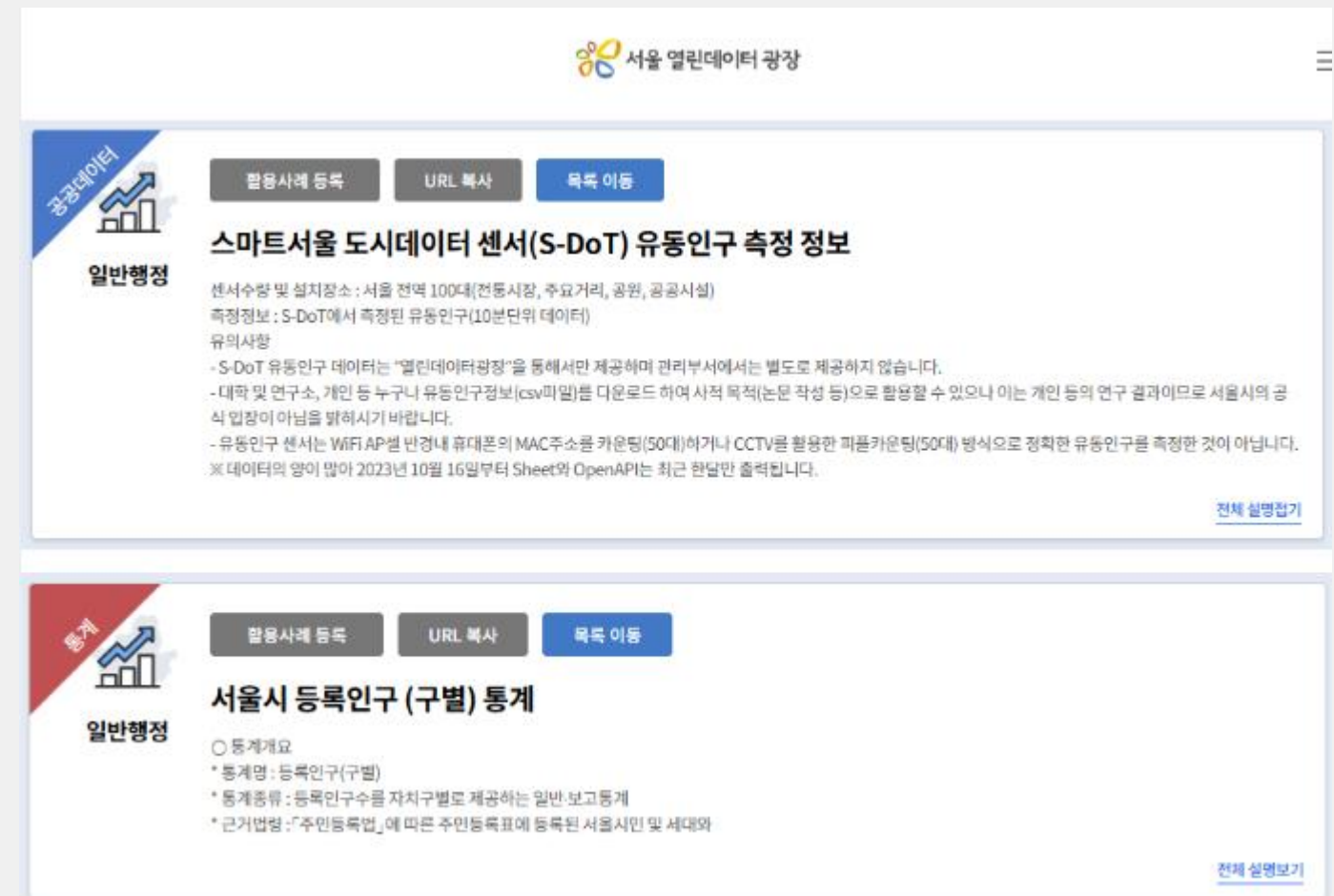
03. 하반기 과정 내용

12월 활동

- Panopticon 시각화를 통한 결론 도출

- 향후 연구 방향

거주 인구, 여행자, 유동 인구 수, 지자체 행정 기관 데이터 등 다양한 변수와 데이터를 추가적으로 고려하여 분석 결과를 보다 정교하게 해석하여 보다 유의미하고 실질적인 결과 도출을 기대함



04. 프로젝트 결과

1. 결과 정리

규모와 진도에 따른 피해정도 종속변수를 가장 잘 예측 하는 XGBoost 모델 생성

→ 피해정도 종속변수와 재난문자 전송빈도 지역별 matplot 비교로 타당성 확보

→ 그 결과, 2022년 ~ 2024년 기준 강원도 동해시와 전라남도 신안군이 가장 피해정도가 컸음

지진 피해정도에 따른 대피소 위치/개수 적합성을 판단

→ 강원 동해시와 전남 신안군이 가장 피해정도가 컸지만 대피소 개수는 많이 부족함. (2022년 ~ 2024년 기준)

→ 그러므로 동해시와 신안군은 지진 대피소 최적 입지 선정 및 마련과 대처방안을 더욱 강구할 필요가 있음.

04. 프로젝트 결과

2. 기대효과

- 1) 강원도 동해시와 전라남도 신안군과 같이 피해가 크지만 대피소가 부족한 지역을 식별하여 대피소 추가 설치를 추진함으로써 **재난 대응력을 강화 가능**
 - ➔ 거주 인구, 유동 인구 수 등 다양한 데이터를 추가로 고려하여 보다 정교하게 해석하고 유의미한 결과를 도출 가능
 - ➔ 동해시의 경우 거주하는 인구가 적어서 대피소가 적을 수 있지만, 여행자가 굉장히 많은 편에 속하는 지역이기 때문에 거주 인구 뿐만 아니라 여행자, 지역별 방문자 수 등을 고려하여 대피소를 추가할 필요가 있음
- 2) 지진 발생 시 안전을 확보하고 **피해를 최소화**하며, 지역사회 재난 **관리의 효율성**을 높이는 데 기여함

05. ALTAIR Tools

1. RapidMiner



RAPIDMINER

- 코드를 작성하지 않고 오퍼레이터 추가 및 조정만으로도 활용 가능
- 데이터 전처리부터 분석과 모델 개발까지 가능
- 분석 결과에 따른 간단한 차트와 시각화 자료 확인 가능

2. Panopticon

Altair Panopticon™

- Altair내의 다양한 데이터 소스와 연결 가능함
- 코드를 작성하지 않고도 직관적인 시각화가 가능함
- 시각화 자료 위에 마우스 포인터를 올리면 이에 해당하는 세부정보를 확인할 수 있음

2024 알테어 동지프 3기 최종 발표회

감사합니다

상명대학교 GRIT 팀

