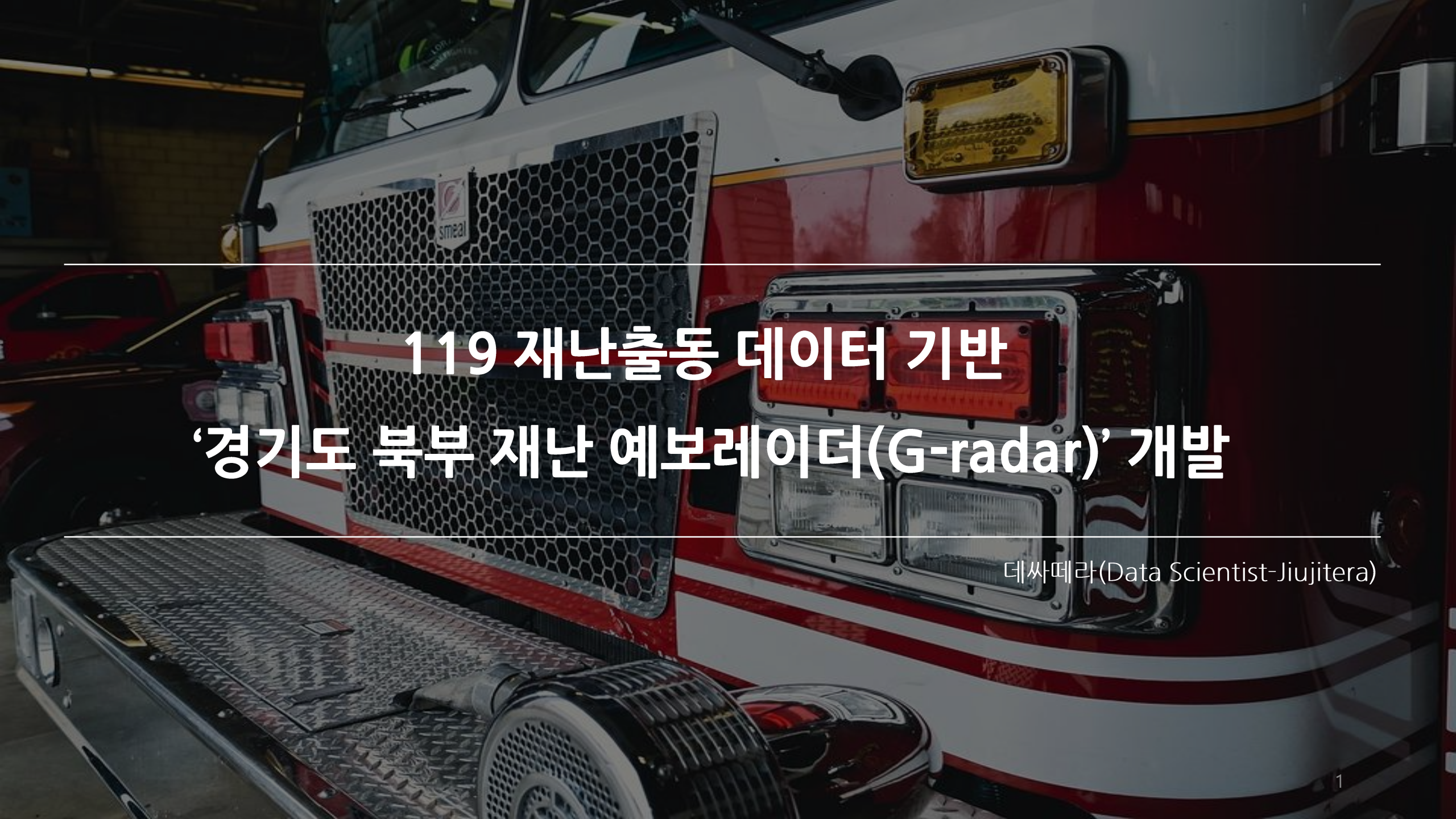


119 재난출동 데이터 기반
‘경기도 북부 재난 예보레이더(G-radar)’ 개발

데싸떼라(Data Scientist-Jiujitera)

Contents

1. 프로젝트 개요
2. 활용 데이터
3. 모델 개발 방법
4. 실험 및 평가
5. 활용 계획 및 기대효과
6. 시연

1. 프로젝트 개요

우리가 해결해야 할 문제는

From. 경기도 북부소방재난본부
북부재난종합지휘센터 손태영님

[What]

“ 경기도 북부지역의 당일 발생 가능한 사건/사고를 지역별로 예측해 주세요 ”

[Why]

“ 반복적으로 발생하며 예상 가능한 재해는 데이터 기반으로 예측하여
사건/사고의 피해를 줄이고 공공의 서비스로 발전시킬 수 있기 때문입니다.”

[How]

“ 기존 119 신고데이터와 사건/사고 발생과 연관 있는 공공 데이터를 활용해
일별/지역별 사건/사고 발생 예측 모델을 개발하고 가시화하여 주세요.”

1. 프로젝트 개요

우리가 집중하고자 하는 문제는

By 데싸떼라

[재난/재해 피해 비용 최소화]

- 사건 발생시 정확한 예측으로 절감되는 비용 > 예측 했지만 사건이 발생하지 않아서 발생하는 비용
- 실제 사건발생시 예측을 정확히 하는 모델 필요

⇒ Recall(재현율)과 F1 Score(재현율과 정밀도의 조화)를 극대화 한 AI 모델 개발

* Recall(재현율): 실제로 일어난 것들 중 일어날 것이라고 예측한 비율
* Precision(정밀도): 일어날 것이라고 예측한 것들 중 실제로 일어난 비율
* F1 Score: 2*(precision*recall)/(precision+recall)

[세부종별 모델 최적화]

- 세부종별 발생 건수, 발생 패턴이 상이한데, 각각의 특성을 충분히 모델에 반영할 수 없을까?

⇒ 세부종별 예측 모델을 별도 개발 후, 발생 예상확률을 한눈에 보기 쉽게 시각화하여 제공

⇒ 전체 사건/사고 중 95.66%를 차지하는 부상, 벌집제거, 교통사고, 자연재해 예측 모델 개발

1위. 부상(86,225건), 2위. 벌집제거(84,037건), 3위. 교통사고(77,449건), 4위. 자연재해(12,398건)

(발생빈도가 적은 세부종은 훈련할 데이터 수가 충분하지 않아 예측 정확도가 높지 않으므로, 모델 개발시 제외)

[시시각각으로 변하는 기상예보를 반영한 예측]

- 기상상황에 많은 영향을 받는 세부종별 사건/사고 발생을 잘 예측하기 위해, 기상 예보를 활용할 수는 없을까?

⇒ 기상청 단기예보 데이터를 모델실행일 전일 기준으로 자동으로 업데이트 해 당일/내일의 사건/사고를 예측

2. 활용 데이터

119 신고 데이터

출동지령일자/출동지령시간/
읍면동/세부종별

- 건수: 271,903건
- 기간: 2016.01.01.~2021.12.31.
- 지역: 250개 읍면동
- 신고종: 4종 (구급, 구조, 기타, 화재)
- 세부종: 11종 (교통사고, 부상, 벌집제거, 자연재해, 산악사고, 수난사고, 대민지원, 산사태, 지원출동(배수), 지원출동(급수), 화재(산불))

재난 발생에 영향을 줄 수 있는 자연적 요인

일별 날씨

평균, 최고, 최저 기온,
평균, 최고 풍속, 강수량

출처: 기상자료개방포털 (https://data.kma.go.kr), 방재기상관측(AWS) 데이터 및 단기예보 데이터

재난 발생에 영향을 줄 수 있는 지역별 요인

지역별(구/군) 인구수

출처: KOSIS 국가통계포털(https://kosis.kr), 주민등록인구현황

재난발생에 영향을 줄 수 있는 시기적 요인

주말 여부

주일比 주말 발생빈도가 높은 예: 부상

출처: 119 신고 데이터 가공

재난발생에 영향을 줄 수 있는 시기적 요인

공휴일 여부

비공휴일 比 공휴일 발생 빈도가 높은 예: 부상

출처: 119 신고 데이터 가공

시기별 · 지역별 · 세부종별 발생 패턴을 반영하는 요인

세부종별 가중치 (weight) : 월&읍면동 기준 사건/사고 평균 발생건수(6개년)

[Weight 계산법]

: Test dataset은 제외하고
매년 월별&읍면동별&세부종별
사건/사고 발생건수를
평균 낸 값으로
weight 계산

Ex) 2월 수내동의
' 벌집제거' 발생 건수 평균(6개년)

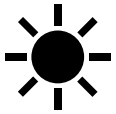
출처: 119 신고 데이터 가공

2. 활용 데이터

독립변수

출동 시점 정보

월(1월~12월),
일(1일~31일)



출동일 기준 읍면동의 날씨 정보

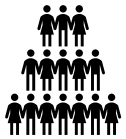
평균/최저/최고 기온,
강수량, 평균/최대 풍속

공휴일 여부

공휴일 or 비공휴일

출동 History

출동월 기준 읍면동의
사건/사고 평균 발생건수(6개년)



출동월 기준 구군별 인구수

읍면동이 속한
구군별 인구수

주말 여부

주말 or 평일

종속변수

- ✓ **벌집제거 신고 발생여부**
- ✓ **부상 신고 발생여부**
- ✓ **교통사고 신고 발생여부**
- ✓ **자연재해 신고 발생여부**

3. 모델 개발 방법

Process	데이터 수집 및 전처리	분류 모델 개발	정확도 검증 및 분류기 결과 시각화
Module	**데이터 수집/분석, 변수 선정** 공공 데이터 (날씨, 인구수) 제공 데이터 가공 데이터 Data * 변수: 시간, 지역, 날씨, 과거 패턴, 휴일 여부, 인구 **분류기가 잘 학습할 수 있도록 데이터 전처리** Cleansing: 결측 데이터가 있는 행 삭제 Encoding: 모델이 학습할 수 있도록 범주형 데이터를 수치형 데이터로 변경 Scaling: 변수들의 각기 다른 스케일 조정	**4대 세부종별 분류기 훈련 및 성능 테스트** • KNN 계열 • Decision Tree 계열 • Random Forest 계열 • Boosting 계열 … * 실험 모델 중 Recall이 높은 상위 3개 분류기 선택 (성능개선을 위해) 상위 3개 분류기의 예측치를 결합하는 **앙상블 모델 생성 및 테스트** Classifier 1 / Classifier 2 / Classifier 3 Voting 최종 예측 * Soft Voting: 상위 분류기의 발생확률을 각각 구한 후 구해진 확률의 평균을 구하여 발생 예측	테스트 데이터로 정확도 검증, 4대 세부종별 예측 모델 생성 최종 모델에 기반해 **4대 세부종별 발생 확률 도출** 단기 기상 예보 분류 모델: 벌집제거 모델 / 부상 모델 / 교통사고 모델 / 자연재해 모델 [예측 결과] ***읍 벌집제거: 78% 부상: 38% 교통사고: 30% 자연재해: 4% 4대 세부종별 발생확률 **시각화**(일별)
Point	사건/사고에 영향을 줄 수 있는 변수 선정	사건/사고를 가장 잘 예측하는 모델 선택 및 결합	Daily 기상 예보 기반 예측

4. 실험 및 평가

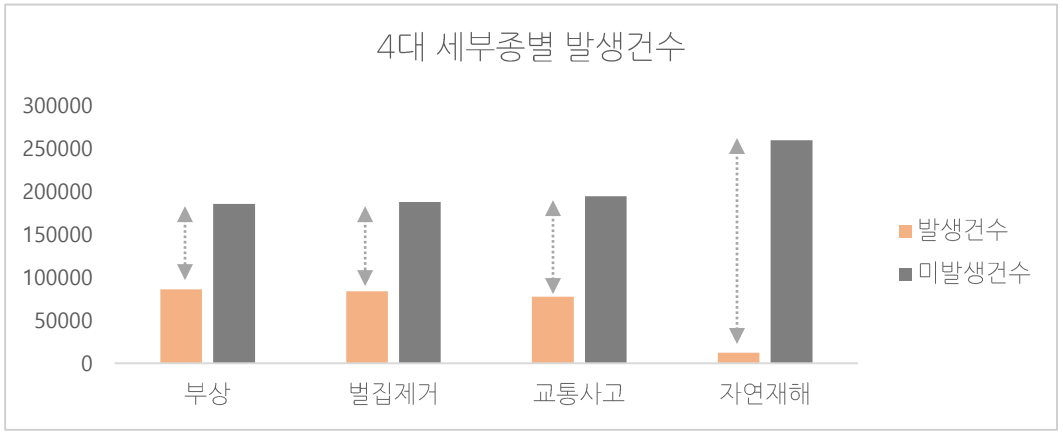
데이터 셋 분리

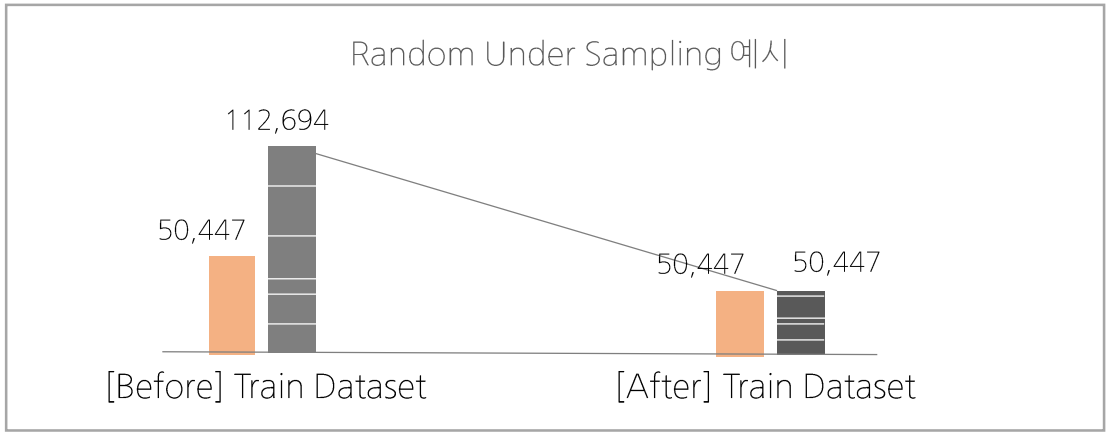
전체 데이터를 Train, Validation, Test 데이터로 분리

70% 모델 훈련용	10% 검증용	20% 최종 평가용

샘플링 작업

예측해야 하는 결과값의 분포가 불균형한 분포를 띄고 있어, 해당 데이터로 학습할 때 발생할 수 있는 문제(과적합 발생 및 재현율 성능 저하)를 개선하기 위해,
데이터의 분포가 높은 값을 낮은 값으로 맞춰주는 Under Sampling 작업 진행





4. 실험 및 평가

벌집제거 모델 평가

[개별 모델 실험결과]

개별 분류기 모델별 성능 평가

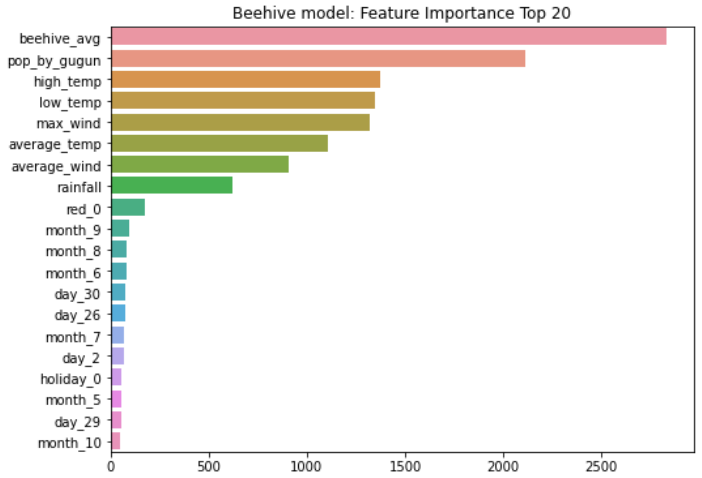
구분	Recall	Accuracy	Precision	F1 Score
Kneighbors Classifier	0.7961	0.7667	0.5882	0.6765
Logistic Regression	0.8488	0.7679	0.5833	0.6915
DecisionTree Classifier	0.7193	0.7457	0.5671	0.6342
RandomForest Classifier	0.8723	0.7700	0.5835	0.6992
GradientBoosting Classifier	0.8619	0.7713	0.5863	0.6979
XGB Classifier	0.8600	0.7825	0.6016	0.7079
LGBM Classifier	0.8592	0.7835	0.6030	0.7087
CatBoost Classifier	0.8597	0.7811	0.5997	0.7065
ExtraTrees Classifier	0.8456	0.7697	0.5862	0.6924

[앙상블 모델 실험결과]

XGB Classifier + LGBM Classifier + CatBoost Classifier
앙상블 모델 성능 평가

구분	Recall	Accuracy	Precision	F1 Score
최종 모델	0.8638	0.7852	0.6091	0.7144

[변수 중요도 결과] (Feature Importance, F1 Score 가장 높은 모델 기준)



- LGBM Classifier

1. 벌집제거 weight
2. 구군별 인구수
3. 최고기온
4. 최저기온
5. 최대풍속
6. 평균기온
7. 평균풍속
8. 강수량
9. 평일 여부
10. 9월 여부

4. 실험 및 평가

부상 모델 평가

[개별 모델 실험결과]

개별 분류기 모델별 성능 평가

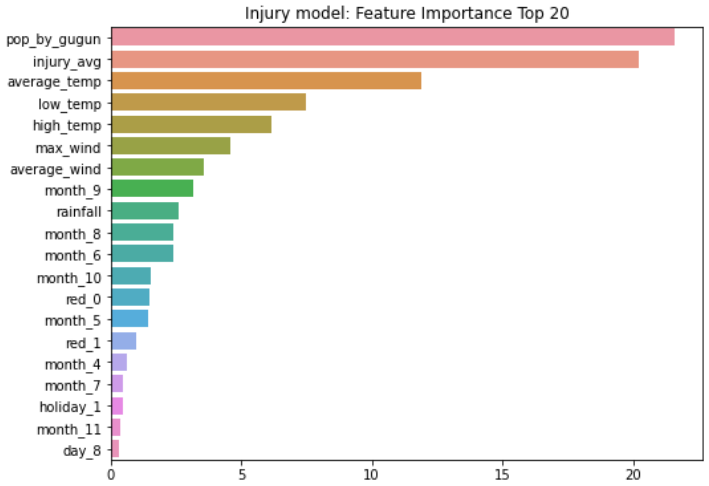
구분	Recall	Accuracy	Precision	F1 Score
Kneighbors Classifier	0.6145	0.6158	0.4236	0.5015
Logistic Regression	0.6843	0.6206	0.4345	0.5315
DecisionTree Classifier	0.5476	0.6057	0.4059	0.4662
RandomForest Classifier	0.7120	0.6329	0.4475	0.5495
GradientBoosting Classifier	0.7078	0.6406	0.4542	0.5533
XGB Classifier	0.7066	0.6465	0.4597	0.5570
LGBM Classifier	0.7140	0.6443	0.4580	0.5581
CatBoost Classifier	0.7192	0.6461	0.4600	0.5611
ExtraTrees Classifier	0.6580	0.6254	0.4366	0.5249

[앙상블 모델 실험결과]

XGB Classifier + LGBM Classifier + CatBoost Classifier
앙상블 모델 성능 평가

구분	Recall	Accuracy	Precision	F1 Score
최종 모델	0.7109	0.6486	0.4639	0.5614

[변수 중요도 결과] (Feature Importance, F1 Score 가장 높은 모델 기준)



- CatBoost Classifier

1. 구군별 인구수
2. 부상 weight
3. 평균기온
4. 최저기온
5. 최대기온
6. 최대풍속
7. 평균풍속
8. 9월 여부
9. 강수량
10. 8월 여부

4. 실험 및 평가

교통사고 모델 평가

[개별 모델 실험결과]

개별 분류기 모델별 성능 평가

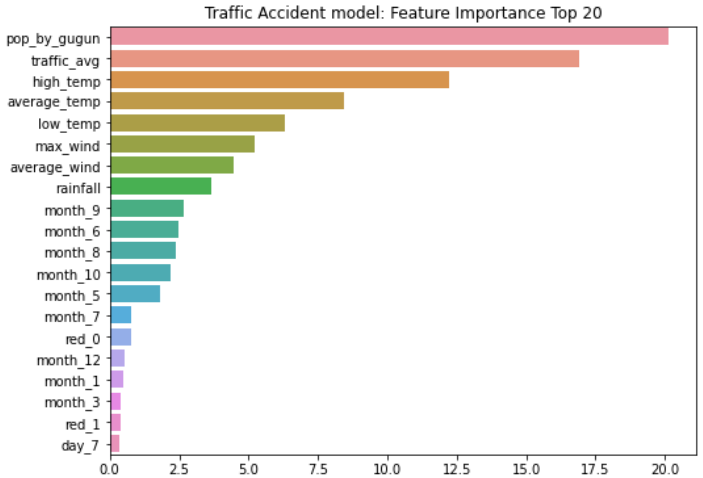
구분	Recall	Accuracy	Precision	F1 Score
Kneighbors Classifier	0.6004	0.6103	0.3871	0.4707
Logistic Regression	0.6909	0.6154	0.4030	0.5090
DecisionTree Classifier	0.5451	0.6064	0.3749	0.4443
RandomForest Classifier	0.7110	0.6219	0.4105	0.5205
GradientBoosting Classifier	0.7041	0.6225	0.4103	0.5185
XGB Classifier	0.7029	0.6279	0.4147	0.5216
LGBM Classifier	0.7109	0.6261	0.4140	0.5232
CatBoost Classifier	0.7176	0.6261	0.4146	0.5256
ExtraTrees Classifier	0.6492	0.6191	0.4012	0.4959

[앙상블 모델 실험결과]

XGB Classifier + LGBM Classifier + CatBoost Classifier
앙상블 모델 성능 평가

구분	Recall	Accuracy	Precision	F1 Score
최종 모델	0.7192	0.6239	0.4075	0.5202

[변수 중요도 결과] (Feature Importance, F1 Score 가장 높은 모델 기준)



- CatBoost Classifier

1. 구군별 인구수
2. 교통사고 weight
3. 최고기온
4. 평균기온
5. 최저기온
6. 최대풍속
7. 평균풍속
8. 강수량
9. 9월 여부
10. 6월 여부

4. 실험 및 평가

자연재해 모델 평가

[개별 모델 실험결과]

개별 분류기 모델별 성능 평가

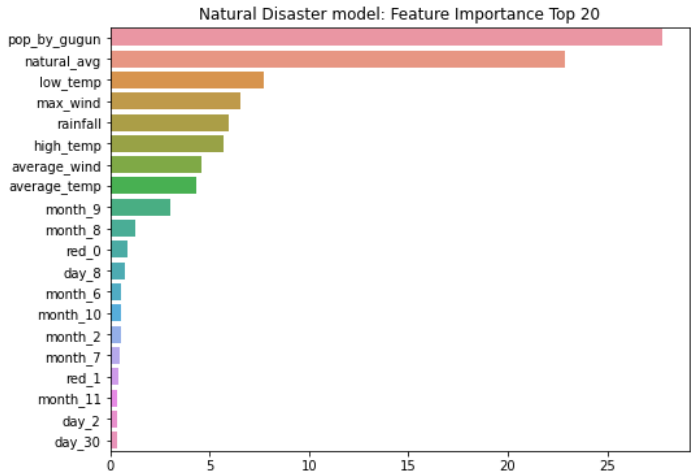
구분	Recall	Accuracy	Precision	F1 Score
Kneighbors Classifier	0.8368	0.8535	0.2185	0.3465
Logistic Regression	0.7734	0.8541	0.2095	0.3297
DecisionTree Classifier	0.8764	0.8579	0.2297	0.3641
RandomForest Classifier	0.8384	0.8965	0.2884	0.4292
GradientBoosting Classifier	0.8542	0.8885	0.2746	0.4157
XGB Classifier	0.9081	0.9058	0.3191	0.4723
LGBM Classifier	0.9002	0.9047	0.3155	0.4672
CatBoost Classifier	0.8986	0.9081	0.3234	0.4757
ExtraTrees Classifier	0.8685	0.9012	0.3031	0.4494

[앙상블 모델 실험결과]

XGB Classifier + LGBM Classifier + CatBoost Classifier
앙상블 모델 성능 평가

구분	Recall	Accuracy	Precision	F1 Score
최종 모델	0.9088	0.9112	0.3324	0.4868

[변수 중요도 결과] (Feature Importance, F1 Score 가장 높은 모델 기준)



- CatBoost Classifier

1. 구군별 인구수
2. 자연재해 weight
3. 최저기온
4. 최대풍속
5. 강수량
6. 최고기온
7. 평균풍속
8. 9월 여부
9. 8월 여부
10. 평일 여부

5. 활용 계획 및 기대효과

AI 모델 작동 과정

✓ Prediction AI model 1: 경기도 북부 사건/사고 예측 모델: 당일의 예측 결과 생성
✓ Prediction AI model 2: 경기도 북부 사건/사고 예측 모델: 다음날의 예측 결과 생성

1. 기상청 단기예보 API를 통해 날씨 데이터를 불러온다.

- AI model 1(당일 예측 결과)의 경우, 전일 기준 오늘의 예보 날씨 데이터를 불러옴
- AI model 2(내일 예측 결과)의 경우, 전일 기준 다음날의 예보 날씨 데이터를 불러옴

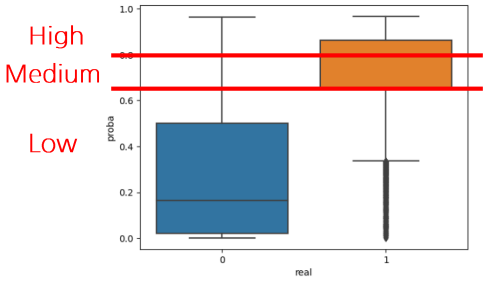
2. 날씨 데이터와 함께 모델에 들어갈 입력 변수들로 데이터프레임을 생성한다.

3. 해당 데이터프레임으로 세부종별 모델에 입력하고 실행한 결과값을 출력 및 시각화한다.

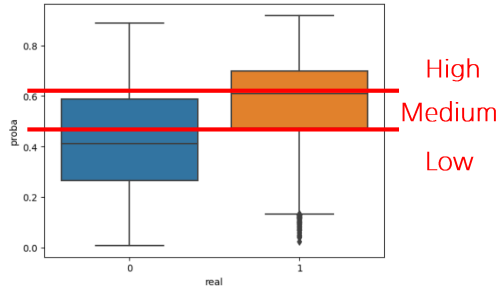
→ 결과1) 읍면동별 세부종별 사건/사고 발생 위험정도에 따른 **지도 시각화**
→ 결과2) 읍면동별 세부종별 사건/사고 발생 위험정도, 예측확률 **표 출력 및 저장**
→ 결과3) 읍면동별 세부종별 사건/사고 발생 위험정도, 예측확률 **검색**

위험정도 측정 기준(저위험~고위험)

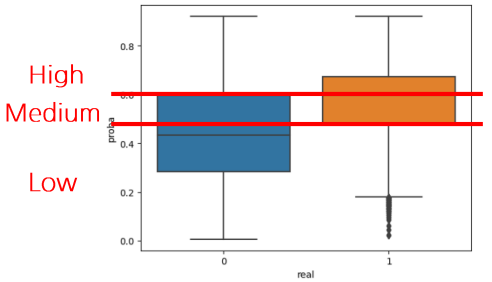
[벌집제거 예측확률 분포]



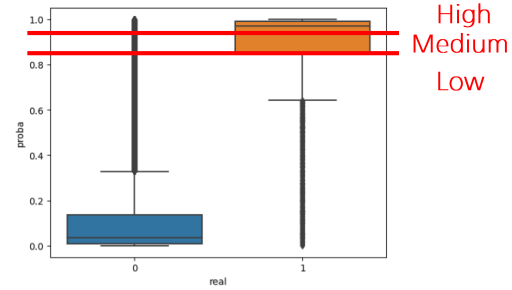
[부상 예측확률 분포]



[교통사고 예측확률 분포]



[자연재해 예측확률 분포]



✓ 실제 신고 발생 데이터의 예측확률 분포 중 중위수가 고위험과 중위험 구분, 1분위(25%)값이 중위험과 저위험 구분 기준이 됨

5. 활용 계획 및 기대효과

결론

✓ 날씨패턴에 영향을 받는 벌집제거는 사건/사고의 발생 위험 정도를 비교적 정확하게 예측할 수 있음

✓ 부상이나 교통사고 등의 사회재난은 실시간으로 변하는 주변 환경 뿐 아니라 예측할 수 없는 인간의 행위로부터 영향을 받기 때문에, 정확한 예측이 어려움

* 부상: [F1 Score] 56.14%
* 교통사고: [F1 Score] 52.02%

✓ 자연재난은 기상상황에 따라 피해로 연계되는 정도가 상이해, 정확한 예측이 어려움

* 자연재난: [F1 Score] 48.68%

→ 사회재난 및 자연재난은 타 데이터와 함께 예측시 활용

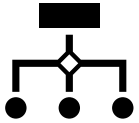
활용 계획

[소방서 현장 활용]
대시보드 및 메일링을 통한
세부종 발생 가능 지역 및
발생 확률 데일리 공유

[공공 서비스와 연계]
예측 정확도가 높은 모델 기반으로
대비를 위한 거주자 대상
재난문자 발송

기대 효과

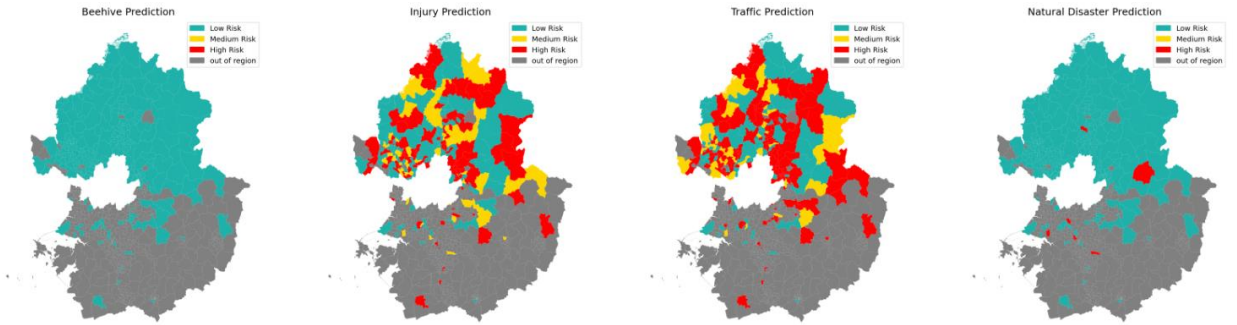
조기 경보 시스템을 활용해
재난 피해 예방 및
사회적 비용 최소화



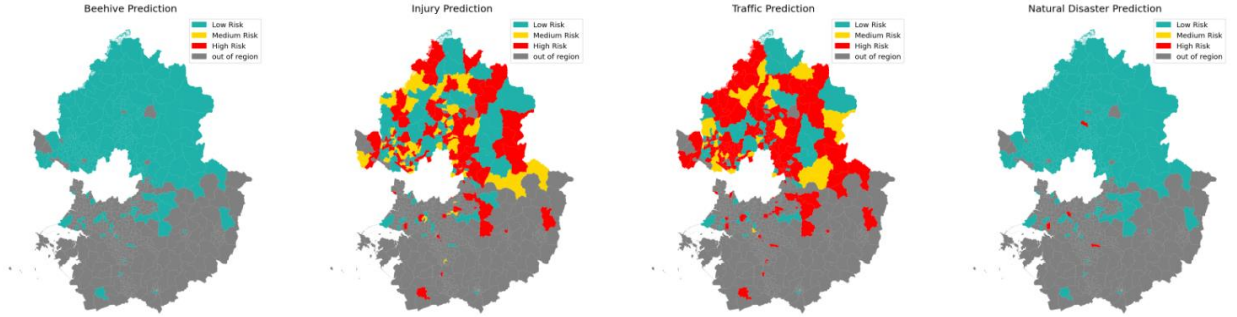
예측 정확도가 벌집제거의 경우,
정확한 예측에 기반해
효율적 인적, 물적 소방자원 배분 가능

5. 활용 계획 및 기대효과

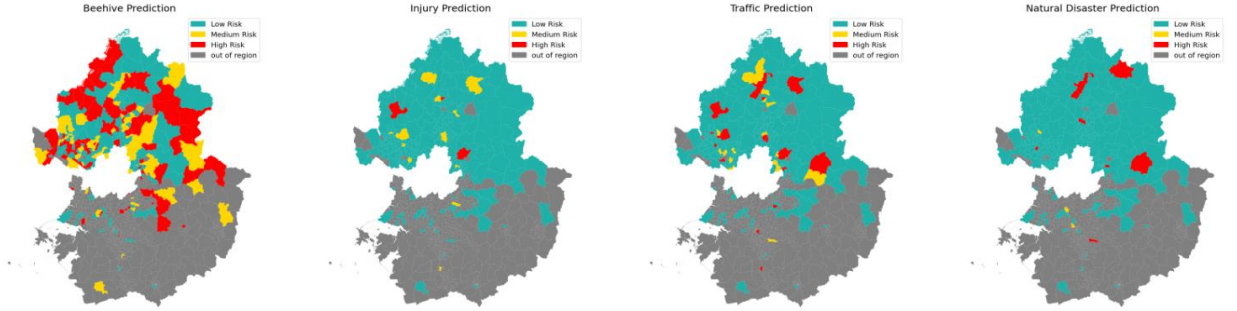
[2022.03.15]



[2022.06.15]



[2022.09.15]



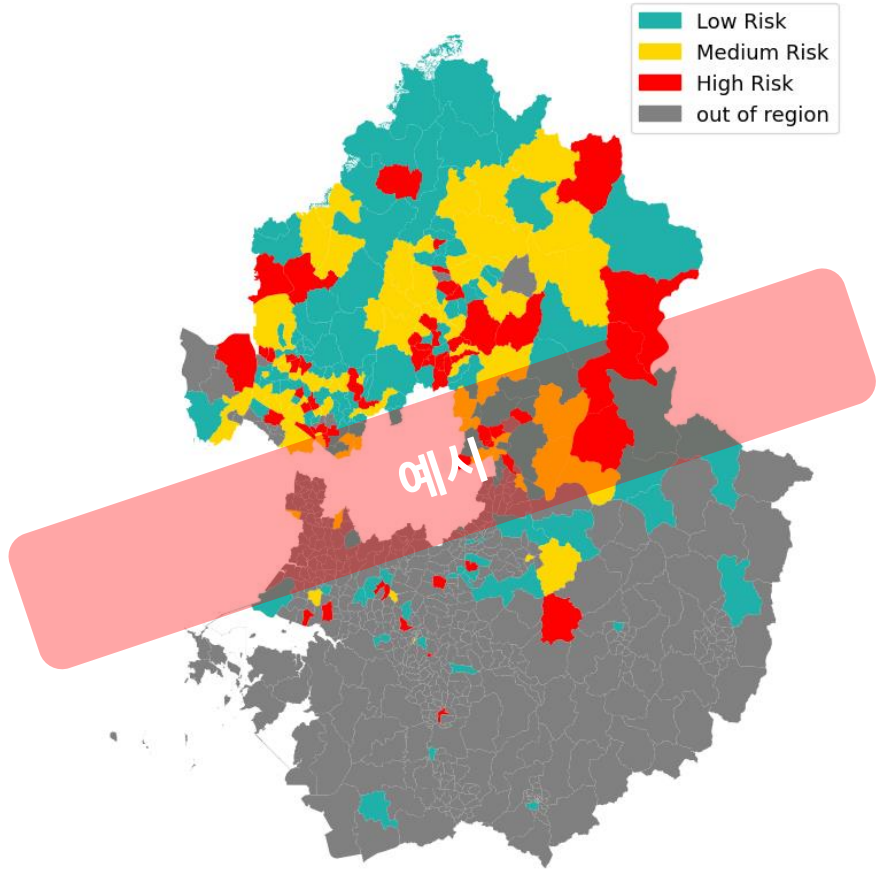
6. 시연

2023년 8월 17일자

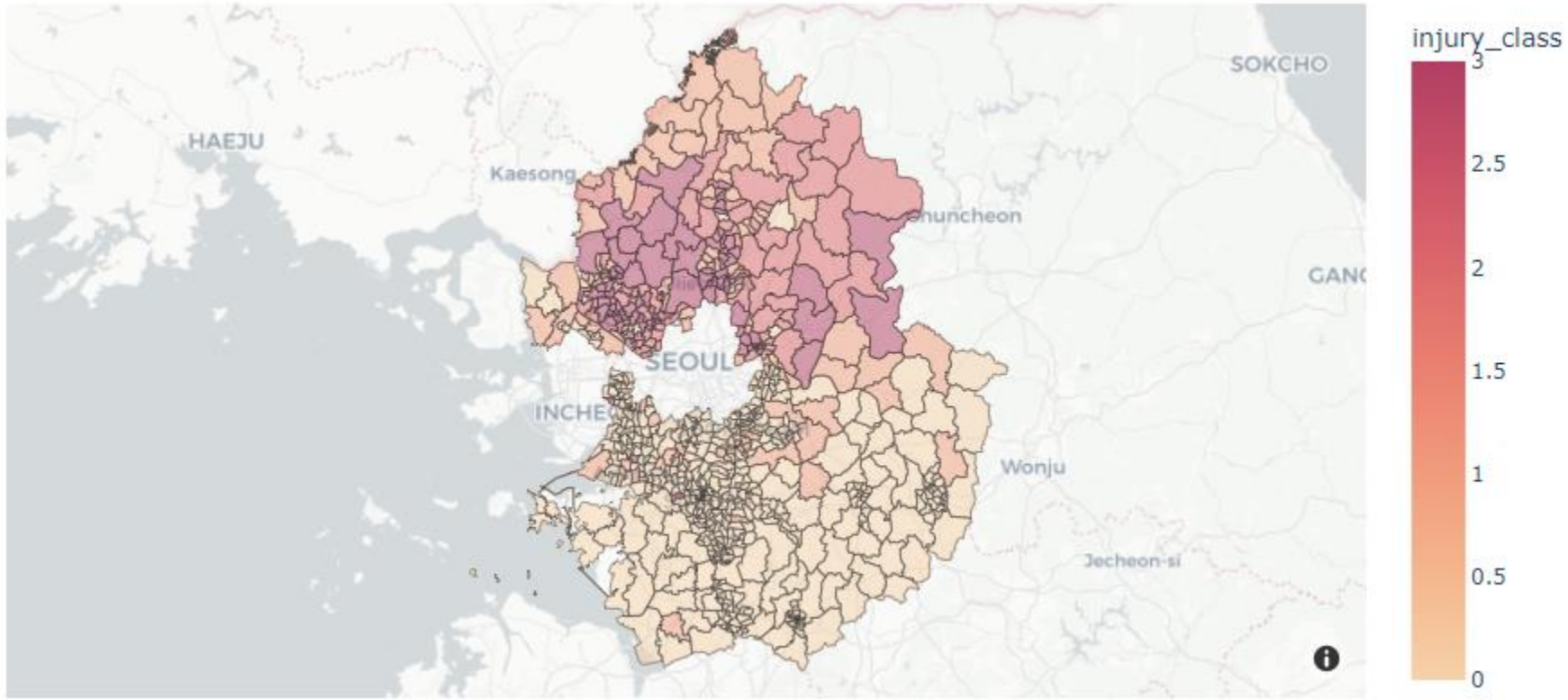
부상 신고발생 예측 ▼

발생 가능 지역(읍면동 기준):
총 19개 지역

자작동	성석동
호평동	주교동
신읍동	토당동
생연동	와부읍
별내면	지금동
의정부동	백석읍
야당동	일산동
신북면	상면
회정동	
오남읍	



6. 시연


HAEJU
Kaesong
SOKCHO
SEOUL
INCHE
Wonju
Jecheon-si
injury_class
3
2.5
2
1.5
1
0.5
0

Q&A

THANK YOU