

2023 서울특별시 빅데이터 캠퍼스 공모전

Made by Festfinder

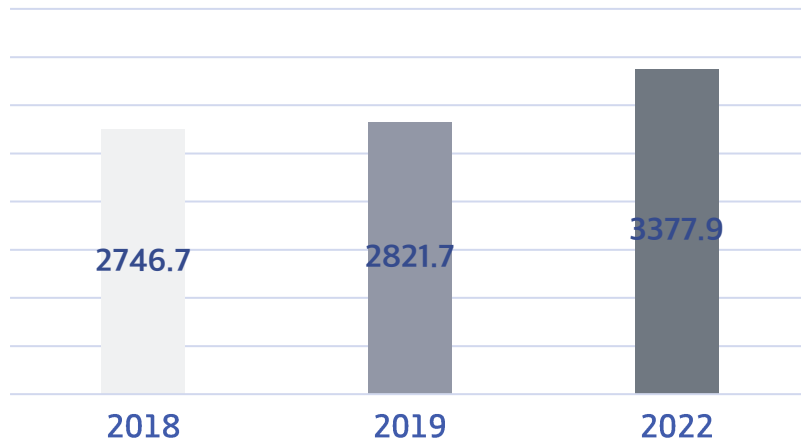
스마트 경로 추천

우선순위 기반 혼잡도 대응 시스템

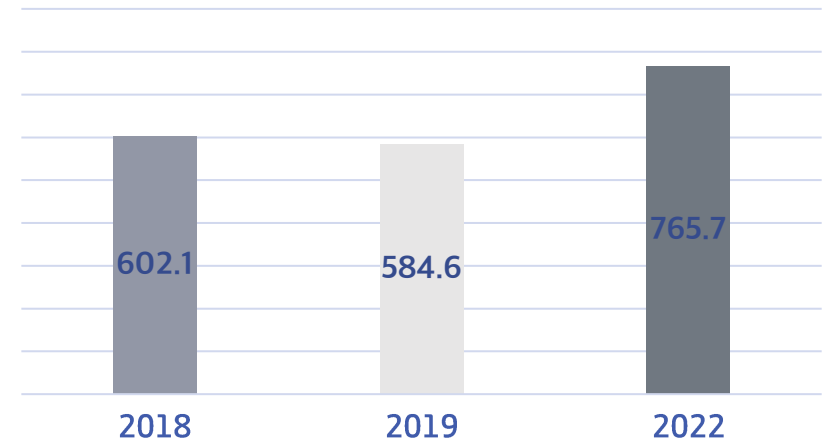
이태원 참사 이후 인파 사고에 대한 관심 증가

- 이태원 참사 이후, 인파 사고는 상당수의 국민들에게 트라우마를 주었음.
- 하지만 2023.03.07 한국관광공사 보도자료를 보면, 축제 기간 총 방문객수와 일평균 방문객수는 여전히 꾸준하게 증가하는 추세를 보임.
- 이러한 방문객 수의 지속적인 증가는 또 다른 인파 사고 발생 가능성을 암시할 수 있음.

축제 기간 총 방문객수
한국관광공사, 보도자료, 2023.03.07

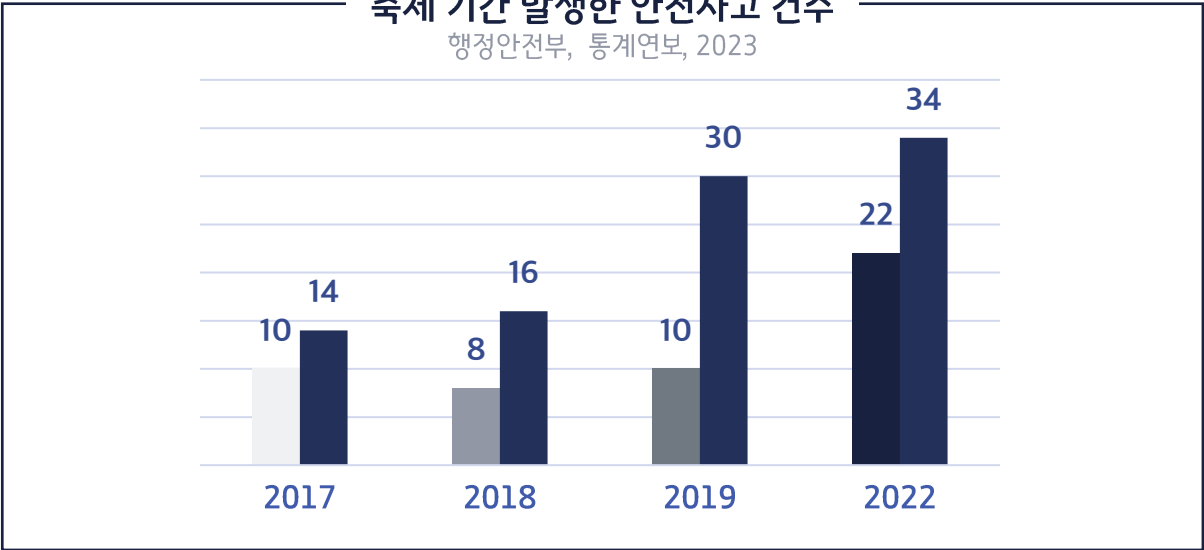


축제 기간 일평균 방문객수
한국관광공사, 보도자료, 2023.03.07



한번 발생하면 큰 사고로 이어지는 인파 사고

- 인파 사고는 사람들이 모여드는 장소나 행사에서 발생하는 대규모 충돌이나 진압, 퇴출로 인해 발생하는 대형사고.
- 이 때문에 축제 사고는 매년 적지 않은 건수의 인파 사고가 발생하고 있음.
- 이러한 사고가 발생하면 ‘나비효과’와 같이 초기의 작은 원인이 많은 사상자를 발생하는 큰 결과를 초래함.



<표 1 - 국내 다중밀집 인파 사고 주요 사례>

일시	장소	사고 내용 및 피해
1992년 2월 8일	서울 올림픽공원 체조경기장 미국 그룹 '뉴키즈 온 더 블록' 공연	수용인원을 초과하여 입장한 관객들이 연쇄적으로 넘어지면서 1명 사망, 50여명 부상
1996년 12월 16일	대구 MBC '별이 빛나는 밤에' 공개방송	관객들이 서로 밀고 당기면서 1명 사망, 5명 부상
1998년 12월 4일	전라남도 순천시 '소년소녀가장돕기 콘서트'	인기 아이돌 그룹 공연 시 학생들이 한꺼번에 몰리면서 10여명 부상
2004년 6월 4일	충청북도 청주시 청주대 개교 기념 음악 공연	공연장 입장 중 인파가 몰리며 뒷사람에게 밀려 13명 부상
2005년 10월 3일	경상북도 상주시 MBC 가요콘서트	관람을 위해 입장하던 시민들이 뒤쪽에서 밀려드는 관객들에게 밀리면서 11명 사망, 110명 부상
2022년 10월 29일	서울 이태원 해밀턴호텔 앞 골목	해밀턴호텔 앞 비좁은 내리막 골목에 인파가 몰리면서 158명 사망, 196명 부상

자료: 뉴시스(2022.10.30.), 백승목(2014.10.17.), Choi, S.-H. et al.(2022.10.30.)로부터 정리

1. 주제 선정 배경

현재 정부의 인파사고 방지 대책

현재 대책의 문제점

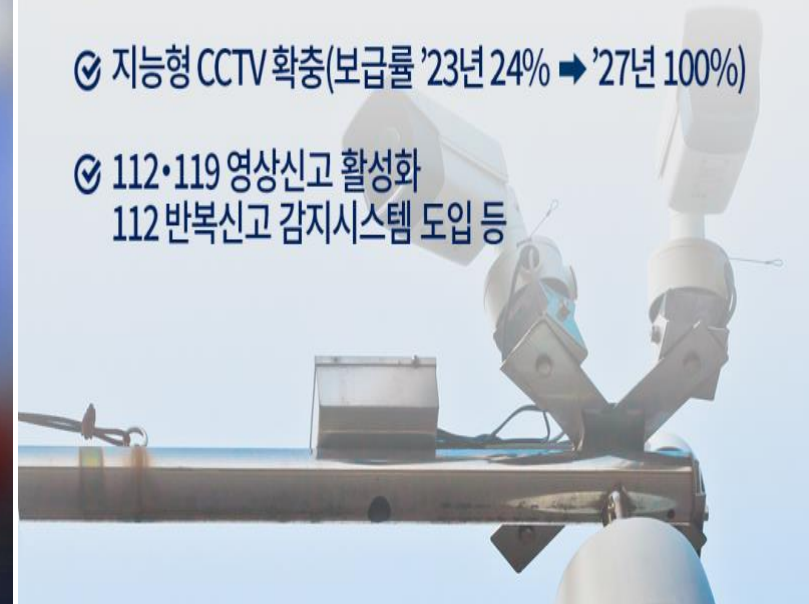
- 실제 현장에 직접 가서 혼잡도 경보를 듣지 않는 한, 지능형 인파감지 CCTV는 사람들에게 해당 장소의 인구 혼잡도를 제공해 줄 수 없음.
- 서울 실시간 도시데이터에서 제공하는 혼잡도 지도는 큰 틀에서의 인구분포만을 나타내므로, 사람들이 실제로 이동하는 도로의 구체적인 혼잡 상황을 알 수는 없음.



위험판단

위험상황 조기파악 체계 구축

- ☑ ICT 기반 현장인파관리시스템 금년 내 도입
- ☑ 지능형 CCTV 확충(보급률 '23년 24% → '27년 100%)
- ☑ 112·119 영상신고 활성화
112 반복신고 감지시스템 도입 등



1. 주제 선정 배경

기존 지도 어플과의 차이점

연구의 필요성 및 목적

- 기존의 지도 앱들은 도보 이동 경로를 추천할 때 대체로 최단 거리나 주요 도로를 우선시하는 방식으로 작동함.
- 이러한 방식은 실제 인구 혼잡도를 고려하지 않기 때문에 사용자들을 동일한 최단 경로로 안내하게 되어, 해당 경로에 혼잡도가 상승할 수 있음.
- 이 문제를 해결하기 위해, 인구 혼잡도를 반영하여 경로를 추천하는 새로운 시스템을 고안.



분석 개요

개발 환경 및 도구

데이터 수집



데이터 분석



지도 관련 시각화



데이터 정제 및 융합

1

도로당 유동인구 매칭

- 출퇴근과 같은 고정적인 혼잡이 아니라 행사와 같은 급격한 인파가 몰리는 혼잡이 위험하므로 주말, 활동시간(8시 - 20시)로 한정된 뒤 도로당 평균 유동인구로 칼럼 매칭 및 전처리

분석용 표준 데이터셋 생성

도로면적

인구 변화율

도로당 유동인구

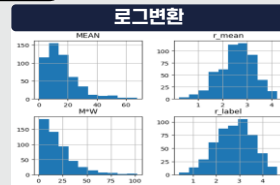
도로 근처 건물 수

도로 근처 버스정거장 수

기상 상태

2

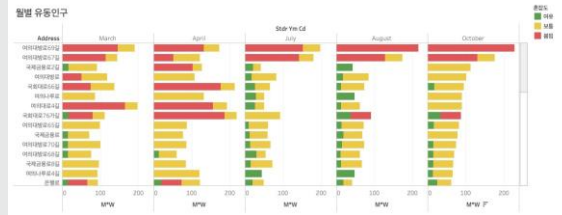
혼잡도 예측을 위한 레이블 생성



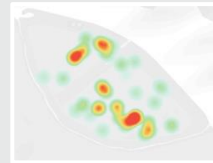
레이블 데이터를 로그 변환을 통해 정규분포에 가깝게 변환하여 새로운 레이블 세트 생성

데이터 분석 및 모델링

유동인구 분석



- 월별 도로의 유동인구를 시각화한 후 유동인구가 많은 지역과 그 지역의 혼잡도를 보여줌



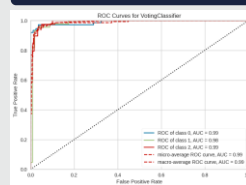
	감소	보통	증가	급증
여유	388	191	125	226
보통	14	63	216	146
붐빔			38	18

- 인구 변화와 혼잡도 간의 관계 파악하여 레이블 검증

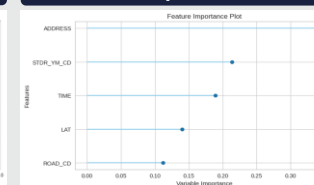
- 시간별 도로 유동인구 혼잡도 시각화

혼잡도 예측 모델링

ROC Curve



Feature Importance 확인



분석 결과 및 인사이트

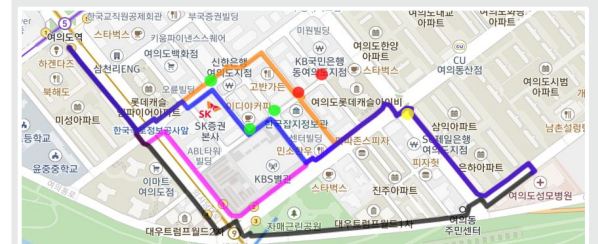
혼잡도 점수 기반의 경로 추천 알고리즘 제안

(1) 경로 종합 평가 점수 산식

$$Score = \sum S_i, S_i = \begin{cases} 10, & \text{혼잡도} = \text{붐빔} \\ 5, & \text{혼잡도} = \text{보통} \\ 1, & \text{혼잡도} = \text{여유} \end{cases}$$

- 경로의 출발지와 도착지의 위도, 경도 좌표를 기반으로 사각형 구역을 정의. 이 구역 내에 포함되는 도로들의 혼잡도를 파악.
- 해당 경로에 포함된 각 도로의 혼잡도 레이블에 따라 가중치를 부여하고, 이 가중치를 합산하여 경로의 총 점수를 계산
- 계산된 총 점수가 가장 낮은 경로를 최적 경로로 선정 (동점 존재할 경우, 거리가 더 짧은 경로를 우선적으로 선택)

(2) 경로 선택 방법



경로	여유	보통	붐빔	총점
경로 1	2	1	0	7
경로 2	1	3	0	16
경로 3	5	3	0	20
경로 4	2	3	1	27

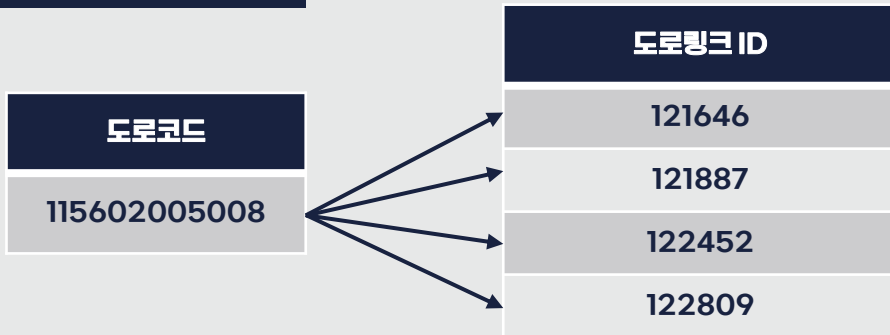
→ 경로 1이 선택됨

2. 분석 방법

데이터 가공

인구 혼잡도 계산 방법

10m 단위 도로링크



각 도로코드는 10m 당 도로구간을 나타내는 도로링크와 1 : M 관계를 가짐.
따라서 도로코드에 속한 도로링크의 수를 통해 해당 도로의 전체 길이 추정 가능.
동일한 도로코드에 속하는 유동인구의 총합을 도로링크 수로 나눔으로써, 각 도로구간의 평균 유동인구. 즉, 도로의 혼잡도를 산출 가능

< 인구 혼잡도 계산 산식 >

$$\text{혼잡도} = \text{도로별 유동인구의 평균} = \frac{\text{도로의 유동인구 합}}{\text{도로링크 수}}$$

인구 변화율 계산 방법

10m도로구간별 유동인구

시간대	1	2	3	4	5	6
시간	0~4시	4~8시	8~12시	12~16시	16~20시	20~24시

유동인구 수가 0인 경우에는 변화율을 계산할 때 분모가 0이 되어 무한대(∞)값을 얻게 되는 문제 발생. 이를 방지하기 위해 계산 전 모든 유동인구 데이터에 1을 더하는 방식으로 0 값을 가진 데이터 포인트에 대해서도 변화율 계산할 수 있도록 함.

비활동 시간대에는 유동인구가 매우 적기 때문에 작은 변화에도 변화율이 크게 나타날 수 있음. 따라서 변화율이 과도하게 반영되는 것을 피하기 위해, 활동 시간대인 3, 4, 5만을 사용하여 인구 변화율을 계산

< 인구 변화율 계산 산식 >

$$\text{인구 변화율} = \frac{\text{현재 시간대의 도로의 유동인구}}{\text{이전 시간대의 도로의 유동인구}} = \begin{cases} \geq 0.7, & \text{급증} \\ \geq 0.1, & \text{증가} \\ \geq 0, & \text{보통} \\ < 0, & \text{감소} \end{cases}$$

2. 분석 방법

데이터 가공

도로면적 파생변수

필요한 칼럼 csv로부터 불러오기

필요한 데이터 컬럼 로드



	도로명	도로코드	도로길이	물리적도로길이	도로폭
4770	선유로40길	4154445	32.0	31.736	2.0
4771	영등포로	3118022	10.0	9.788	2.0
4772	양평로	3118017	25.0	25.313	2.0

도로명, 도로코드, 도로 길이, 물리적 도로길이, 도로 폭

파생변수 생성

1. 도로 면적 계산 산식

```
rd_df['도로면적'] = rd_df['물리적 도로길이'] * rd_df['도로 폭']
```

2. 도로코드를 기준으로 데이터 그룹화한 후, 각 도로코드에 해당하는 면적 집계

	도로명	도로코드	물리적 도로길이	도로 폭	도로면적
4770	선유로40길	4154445	31.736	2.0	63.472
4771	영등포로	3118022	9.788	2.0	19.576
4772	양평로	3118017	25.313	2.0	50.626
4773	영신로37길	4154773	21.914	2.0	43.828
4774	가마산로	3000023	70.198	2.0	140.396

➡

		물리적 도로길이		도로 폭		도로면적	
	도로명	도로코드	sum	count	sum	count	sum
63로	가마산로	3118001.0	594.283	1	25.0	1	14857.075
3000023.0	가마산로	3000023.0	6352.265	56	245.0	56	99389.426
4154002.0	가마산로31길	4154002.0	679.633	4	40.0	4	6796.330
4154003.0	가마산로32가길	4154003.0	167.170	1	2.0	1	334.340
4154004.0	가마산로32길	4154004.0	185.792	1	12.0	1	2229.504

도로 주변 건물수 파생변수

필요한 칼럼 shp로부터 불러오기

필요한 데이터 컬럼 로드



행정동코드			도로코드		
	행정동코드	도로코드		행정동코드	도로코드
0	11560660	4154335	618340	11560710	3116010
1	11260565	4118020	618387	11560710	4154151
2	11260590	4118396	618388	11560515	4154059

행정동코드, 도로코드

파생변수 생성

1. 도로 주변 건물 수 계산 방법

도로코드를 기준으로 데이터를 그룹화한 후, 같은 도로코드를 가진 건물 수 집계

2. 결과

행정동코드	도로코드	행정동코드	도로코드
0	11560660	4154335	618340
1	11260565	4118020	618387
2	11260590	4118396	618388

행정동코드	도로코드	행정동코드	도로코드
11560710	3116010	11560710	4154151
11560515	4154059		

주변건물수		주변건물수	
도로코드	도로코드	도로코드	도로코드
2005008	129	4154818	5
2005010	46	4154826	3
2118001	18	4154835	2
2118002	53	4154836	6
3000023	289	4857947	1

2. 분석 방법

데이터 가공

도로 주변 버스정류장 수 파생변수

필요한 칼럼 csv로부터 불러오기

필요한 데이터 컬럼 로드



전국 버스정류장
위치정보.csv

	정류장번호	정류장명	위도	경도
55044	GGB118000027	경방타임스퀘어.신세계백화점	37.517550	126.906233
55045	GGB118000029	영등포역	37.517117	126.909133
55046	GGB118000032	경방타임스퀘어.신세계백화점	37.518417	126.905433

정류장번호, 정류장명, 위도, 경도

파생변수 생성

1. 도로 주변 버스정거장 수 계산 방법

위도와 경도 기반으로 반경 200미터 내의 버스 정류장 수 계산

2. 결과



전국 버스정류장
위치정보.csv

+



서울시 10m단위
도로구간
공간데이터.csv


→

	도로코드	위도	경도	버스정류장수
0	2005008	37.531846	126.881940	1
1	2118001	37.527963	126.928355	3
2	2118002	37.521731	126.924333	3
3	2121004	37.566270	127.147512	0
4	3000028	37.460288	126.695246	0

기상상태 파생변수

필요한 칼럼 csv로부터 불러오기

필요한 데이터 컬럼 로드



조건별 기후
통계분석.csv

여의도 지역으로 필터링

	일시	평균기온 (°C)	최고기온 평균 (°C)	최저기온 평균 (°C)	강수량 (mm)	평균풍속 (m/s)	최대풍속 (m/s)	평균현지기압 (hPa)	최고현지기압 (hPa)	최저현지기압 (hPa)
0	1.0	-0.9	3.6	-4.8	11.3	2.6	7.7	1013.8	1025.7	1000.2
1	2.0	1.0	6.0	-2.9	22.7	2.9	8.5	1011.4	1024.0	1000.8
2	3.0	6.3	12.4	1.0	9.6	3.1	9.1	1010.7	1025.1	998.4

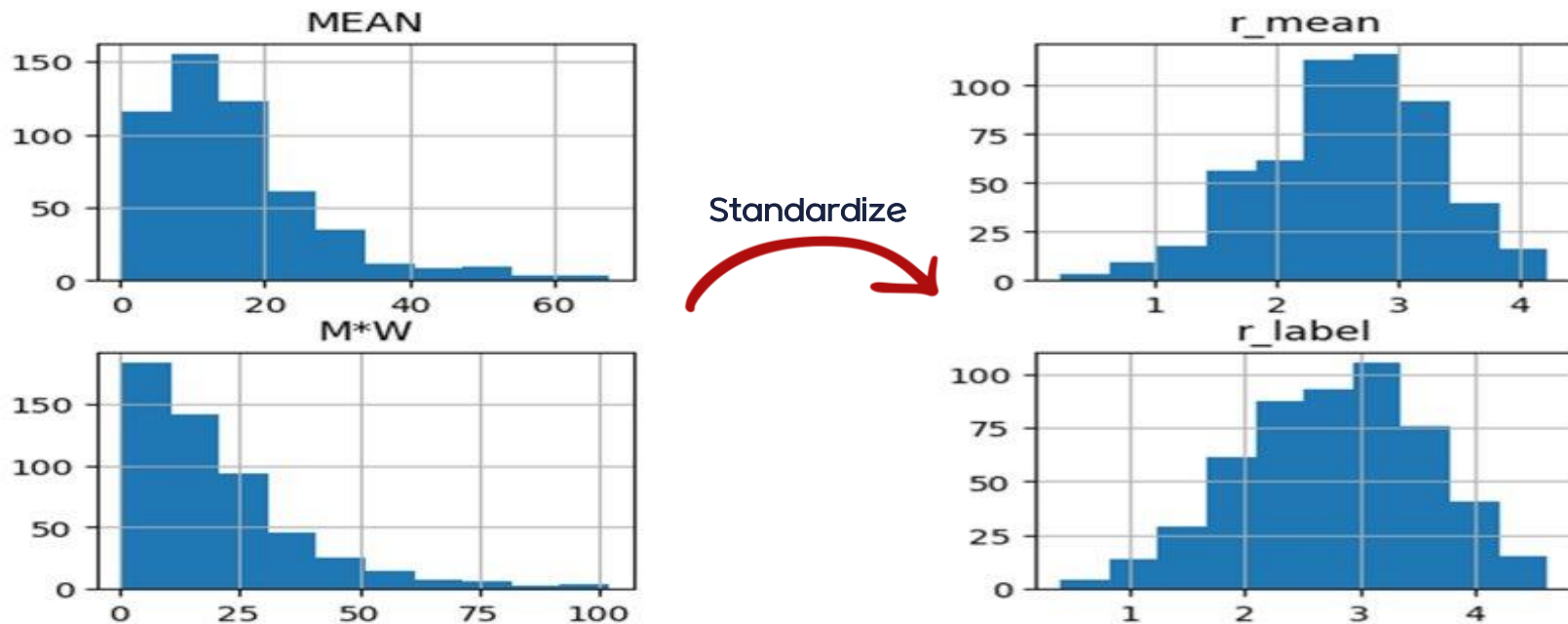
	일시	최저현지기압 (hPa)	평균해면기압 (hPa)	최고해면기압 (hPa)	최저해면기압 (hPa)	평균상대습도 (%)	최소상대습도 (%)	합계일조시간 (hr)	합계전천일사량 (MJ/m²)
0	1.0	1000.2	1024.7	1036.9	1010.8	56.0	22.0	193.3	245.05
1	2.0	1000.8	1022.2	1035.3	1011.3	59.0	17.0	180.0	267.04
2	3.0	998.4	1021.3	1035.8	1008.7	45.0	9.0	270.8	458.62

평균기온(°C), 최고기온 평균, 최저기온 평균, 강수량, 평균풍속, 최대풍속, 평균 현지 기압, 최고 현지 기압, 최저 현지 기압, 평균 해면 기압, 최고 해면 기압, 최저 해면 기압, 평균 상대 습도, 최소 상대 습도, 합계 일조시간, 합계 전천 일사량

2. 분석 방법

데이터 레이블링

로그 변환을 통한 정규화



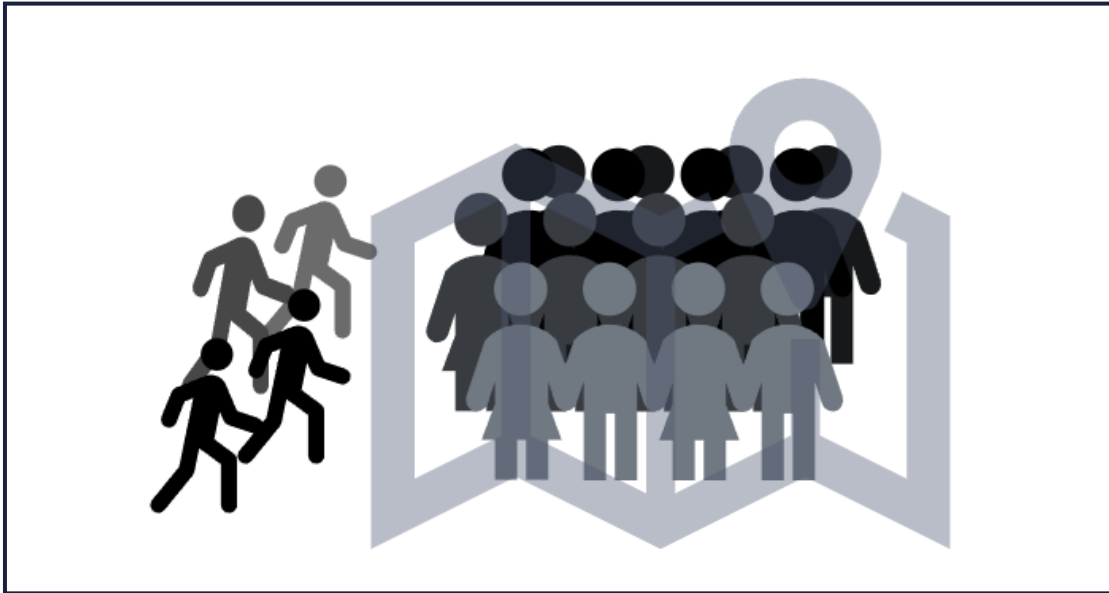
혼잡도 예측을 위한 모델링 접근 방법

- ‘혼잡’이라는 용어에 대한 명확한 정의 수립
→ 단순히 유동인구가 많은 지역을 의미하는 것이 아니라, 유동 인구가 많고 지속적으로 증가하는 추세를 보이는 지역을 **혼잡한 지역**으로 간주.
- 혼잡도가 위험한 지역을 선정하는 방법 = 유동인구 × 인구 변화율
- 유동인구 데이터의 **왜도가 높기 때문에**, 데이터를 **로그 변환**하여 정규 분포와 유사한 형태로 조정한 뒤, 레이블링 하여 분류 모델링에 적용함.
- 다른 변수들에 대해서는 **최소-최대 스케일링(min-max scaling)**을 통해 값의 범위를 조정하고 모델링에 적합하게 만들.

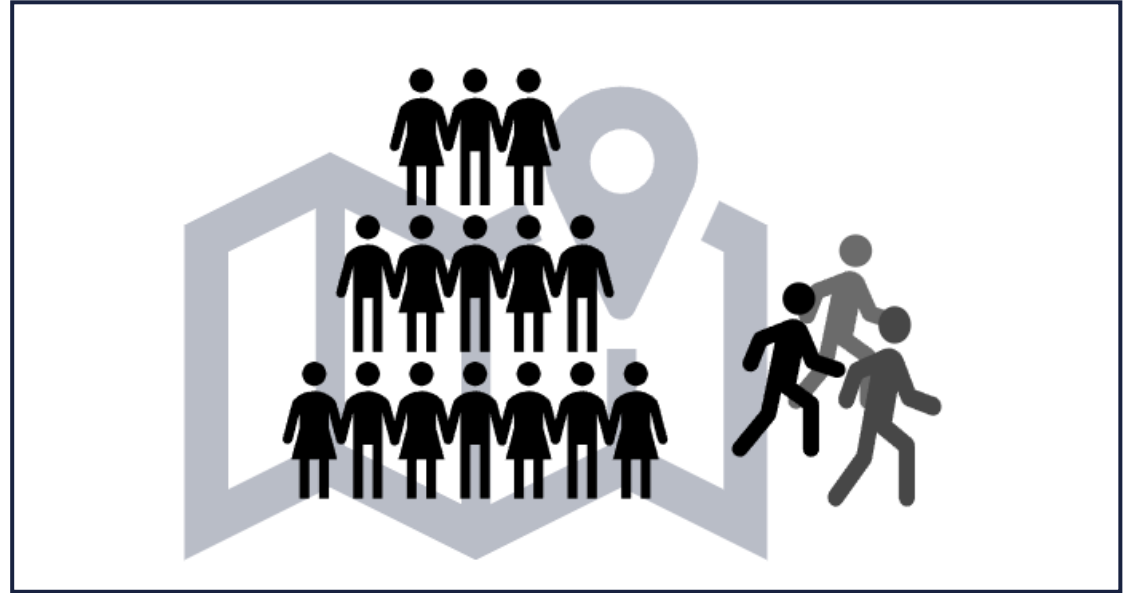
2. 분석 방법

유동인구 변화 가중치 설정

유동인구가 급격히 증가하는 곳에 가중치를 부여



유동인구가 이탈 중인 곳은 분석에서 배제 (낮은 중요도)



각 지역의 유동 인구 증가율을 시간별로 계산하여, 증가율이 **3분위수 (Q3)**를 초과하는 지역을 ‘**급증**’ 지역으로 정의하고 이들에게는 **높은 가중치인 2**를 부여함.

평균 증가율 이상을 보이는 지역에는 **가중치 1**을, **평균 이하 증가** 혹은 **감소**하는 지역에는 각각 **가중치 1**과 **0.7**를 적용하여, 유동 인구의 감소가 분석의 중요성을 감소시키므로 이들 지역은 **분석에서 제외** 혹은 **중요도를 낮춤**.

이러한 방법으로 조정된 유동 인구 데이터를 사용함으로써, 급격한 인구 증가를 경험하는 지역은 분석에서 더욱 부각되고, 인구가 줄어드는 지역은 중요도가 낮아지게 됨.

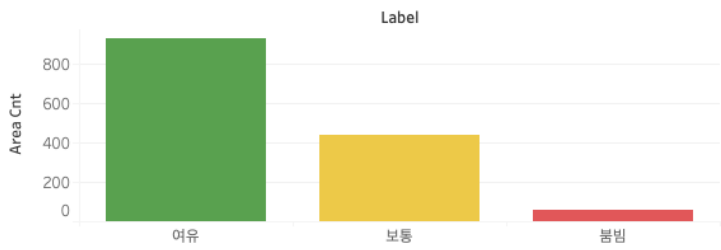
2. 분석 방법

유동인구 분석

혼잡도와 시간대 인구변화 관계

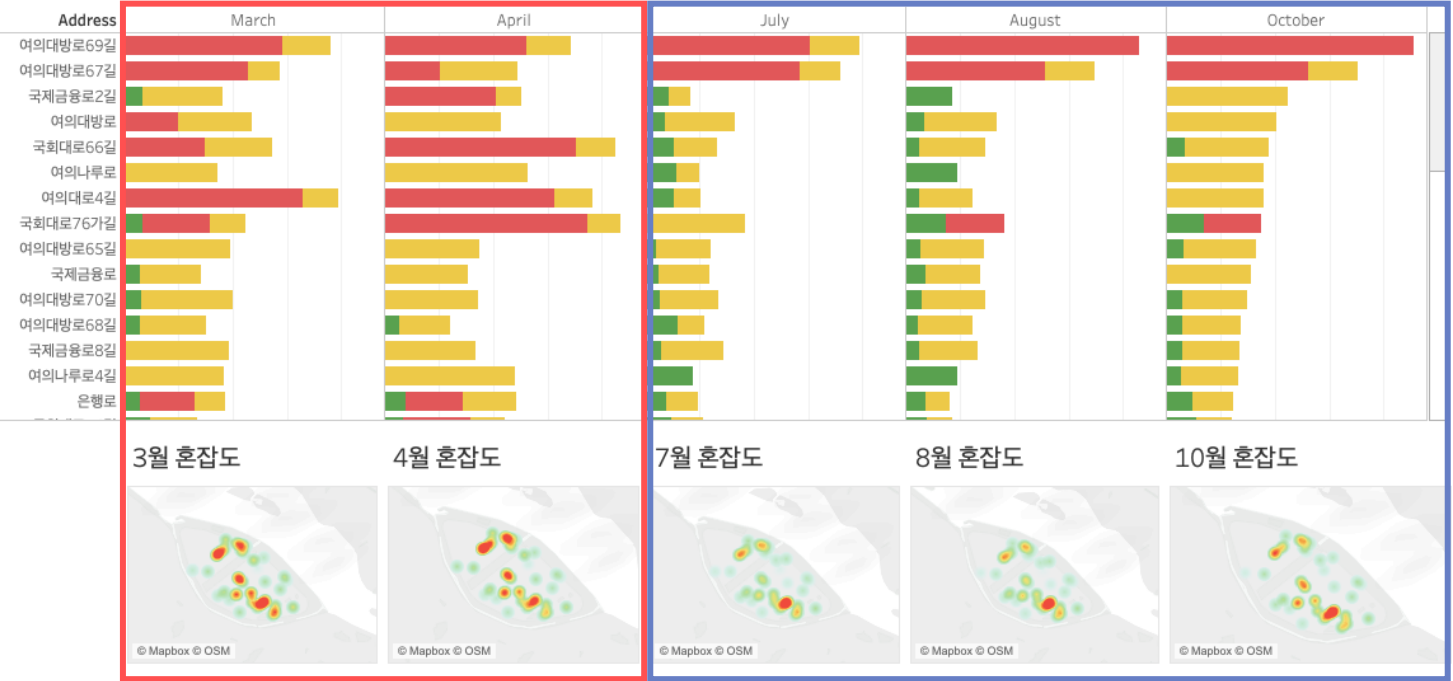
Label	변화			
	감소	보통	증가	급증
여유	388	191	125	226
보통	14	63	216	146
붐빔			38	18

혼잡도 분포



- 혼잡도가 ‘여유’ : 인구 변화가 ‘감소’하거나 ‘급증’하는 경우가 많이 나타남. 이는 여유로운 상태의 지역이나 시간에 인구 변동성이 크다는 것을 의미
- 혼잡도가 ‘보통’ : 인구 변화가 ‘증가’하거나 ‘급증’하는 경우가 다른 상태보다 상대적으로 높은 것으로 나타남. 이는 ‘보통’ 상태가 인구 증가로 이어질 가능성이 높음을 시사함.
- 혼잡도가 ‘붐빔’ : ‘감소’나 ‘보통’ 상태의 인구 변화는 관찰되지 않았으며, 오로지 인구 ‘증가’와 ‘급증’만이 존재함. 이는 매우 혼잡한 상태가 이미 인구가 많은 상황에서 발생하며, 더욱 더 많은 인구 증가를 가져올 수 있음.

월별 유동인구



3월과 4월의 유동 인구 분포가 7월, 8월, 10월의 분포와 뚜렷이 구분됨

해석 1.

유동 인구에 계절적 변동성이 존재한다는 것을 상당한 정도로 추론 가능

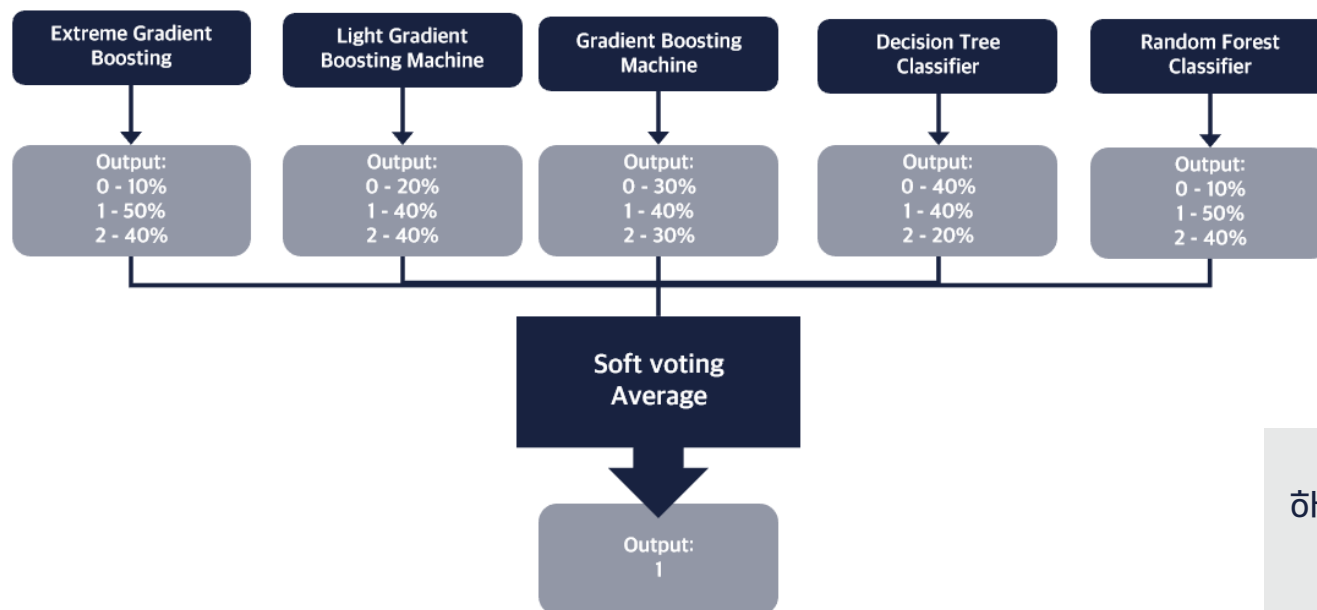
해석 2.

5월에 발생한 메르스와 같은 전염병의 영향으로 유동인구가 감소된 것으로 추론 가능

2. 분석 방법

데이터 분석 및 모델링

- 앙상블 모델링 (Voting)



하이퍼 파라미터 튜닝을 한 개별 모델을 결합하여 최종 예측을 수행하는 앙상블 기법을 사용하여 예측의 정확성을 향상시키고자 함

장점 1. 예측 정확성 향상

- 모델 결합**: 다양한 모델은 다양한 특징과 패턴을 학습하며, 이는 전체 앙상블의 예측 성능을 향상시킴
- Robust한 예측**: 앙상블 모델은 이상치에 민감하지 않으며 안정적인 예측이 가능

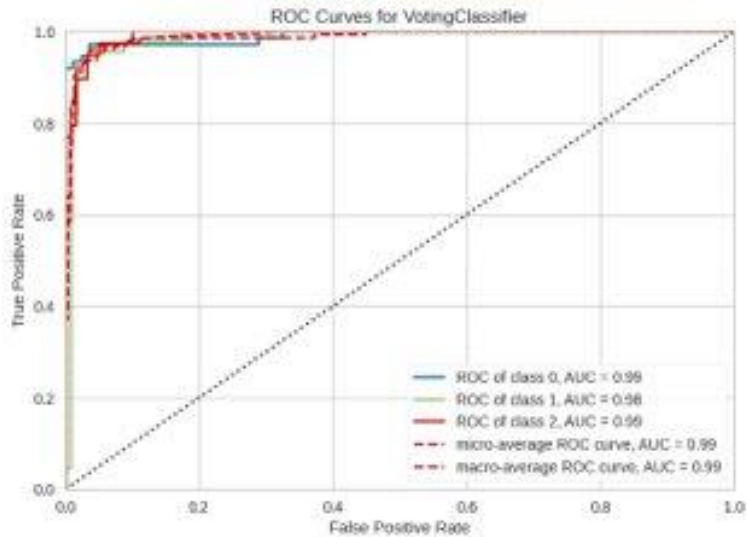
장점 2. 모델의 일반화 성능 개선

- 과적합 방지**: 개별 모델들이 서로 오차를 상쇄하여 모델의 일반화 성능 향상

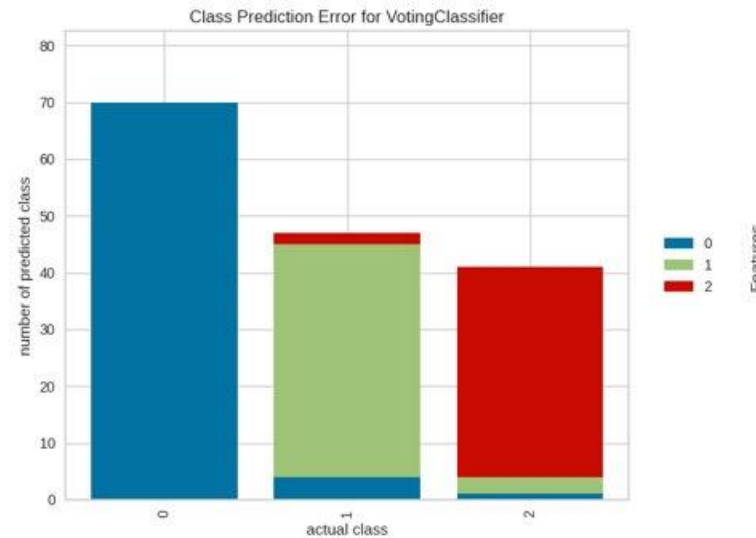
3. 분석 결과

분석 결과

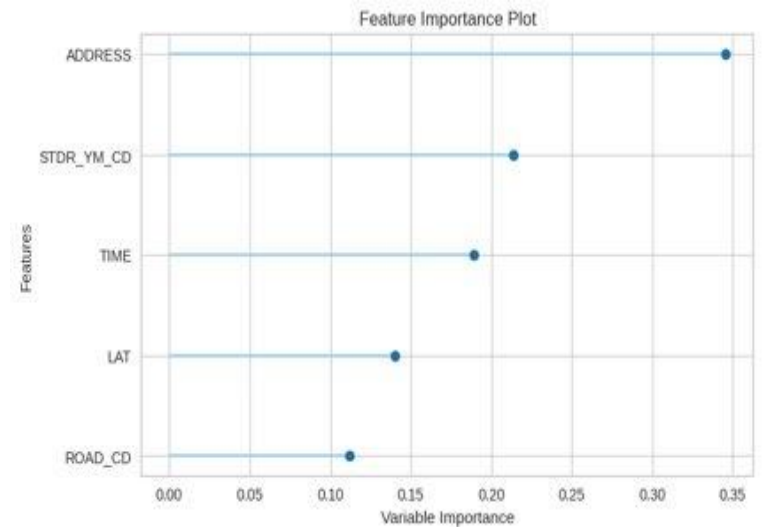
ROC Curve



각 클래스에 대한 모델의 예측 오류



Feature Importance 확인



몇 개의 파생변수 만으로 충분한 예측력을 보임.

위치와 시간적 요인이 혼잡도 예측에 **상당한 영향**을 미친다는 것을 확인할 수 있었음.

분석은 월별 데이터를 기준으로 하고 여의도 지역에 한정되어 진행되었으며, 이는 결과에 영향을 미쳤을 것으로 보임.

장기간에 걸쳐 축적된 유동 인구 데이터, 날씨 정보, 지역 특성 등을 포함한 종합적인 데이터를 활용하여 모델을 학습시킬 경우, 더욱 의미 있는 분석 결과를 도출할 수 있을 것으로 기대됨.

3. 분석 결과

활용방안 1 - 혼잡도 점수 기반의 경로 추천 알고리즘 제안

• 경로 종합 평가 점수 산식

Score = \sum S_i, S_i = \begin{cases} 10, & \text{혼잡도} = \text{불뺨} \\ 5, & \text{혼잡도} = \text{보통} \\ 1, & \text{혼잡도} = \text{여유} \end{cases}

• 경로 종합 평가 점수 산출 방법론

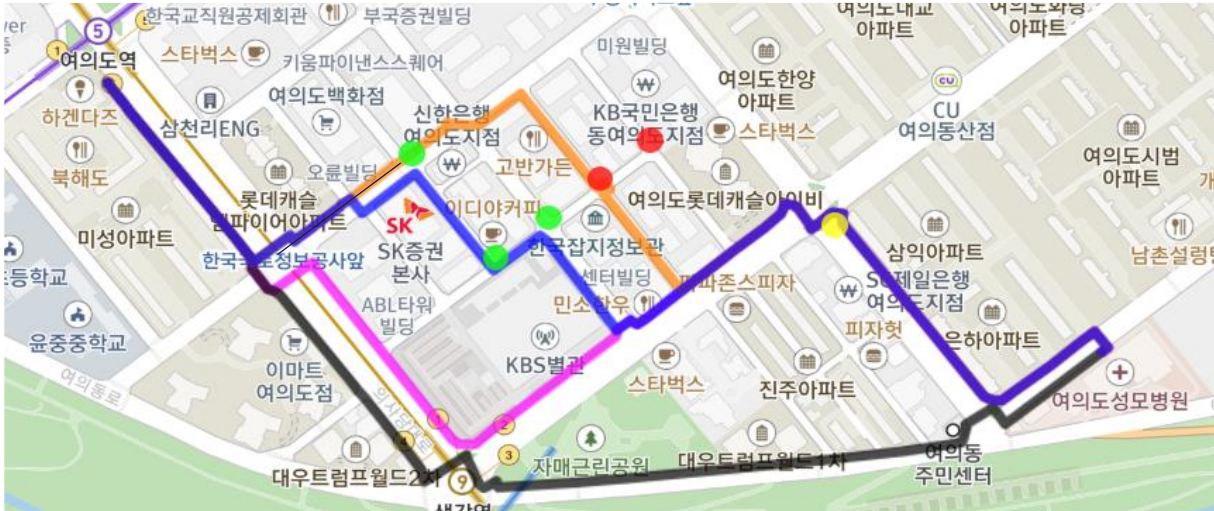
- 1. 경로의 출발지와 도착지의 위도, 경도 좌표를 기반으로 사각형 구역을 정의. 이 구역 내에 포함되는 도로들의 혼잡도를 파악함.
- 2. 해당 경로에 포함된 각 도로의 혼잡도 레이블에 따라 가중치를 부여하고, 이 가중치를 합산하여 경로의 총 점수를 계산.
- 3. 계산된 총 점수가 가장 낮은 경로를 최적 경로로 선정. (동점 존재할 경우, 거리가 더 짧은 경로를 우선적으로 선택)

• 경로 생성 방법론

T-MAP API를 활용하여 보행자 이동 동선 생성

⇒ 출발지에서 목적지까지의 다중 경로 생성 후 앞서 정의한 경로 종합 평가 점수를 이용해 경로 평가

• 경로 종합 평가 점수 산출 예시



빨간색 = 불뺨
노란색 = 보통
초록색 = 여유

경로	여유	보통	불뺨	총점
경로 1	2	1	0	7
경로 2	1	3	0	16
경로 3	5	3	0	20
경로 4	2	3	1	27



경로 1의 점수가 가장 낮으며 거리도 가장 짧으므로 경로 1을 선택

프로젝트 기대효과



이용자 측면

- **개인의 안전 향상** : 경로 추천을 통해 유동인구가 많은 지역을 피함으로써, 사고 위험이나 전염병 노출의 위험 감소
- **시간 절약 및 편의성 증대** : 사용자는 덜 혼잡한 경로를 선택함으로써 통행 시간을 단축하고 더 쾌적한 이동을 할 수 있음



행정 기관 측면

- **사고 예방을 위한 사전 조치 가능** : 혼잡도 예측 모델을 통해 사고 발생 가능성이 높은 지역을 사전에 파악하고, 필요한 조치를 취할 수 있음.
- **교통 혼잡 완화** : 혼잡한 경로를 피하도록 유도함으로써 교통 흐름을 개선하고 혼잡을 완화할 수 있음
- **공공 안전 강화** : 혼잡도 정보를 활용하여 공공의 안전을 유지하고 강화하는데 기여할 수 있음.



행사 기획 측면

- **행사 기획 및 배치 최적화** : 혼잡도 예측 모델을 활용하여 행사의 위치 선정, 인원 배치, 물류 계획 등을 보다 효과적으로 수립할 수 있음.
- **안전 사고 예방** : 혼잡도 예측 결과를 바탕으로 동선 확보를 하여 안전사고를 예방할 수 있게 됨.

한계점

컴퓨팅 자원의 한계

사용 가능한 컴퓨팅 파워의 한계로 인해, 데이터를 축소하게 되었음.

- **파생 변수의 고려 부족** : 유동 인구 예측에 필요한 다양한 파생 변수들이 존재하지만, 이러한 모든 변수들을 충분히 고려하지 못함.
- **시계열 데이터의 부족** : 2015년의 데이터만을 사용하여 장기적인 패턴과 추세를 파악하기에는 불충분했음.
- **지역적 범위의 제한** : 분석이 여의도 지역으로 한정되어 있어, 다른 지역의 유동인구 변화와 비교하거나 일반화하기 어려움.

데이터 자체의 아쉬움

- **공간 데이터 매핑의 정확성 문제** : 위도와 경도 데이터의 변환 과정에서 정확한 위치 매핑에 실패하여, LineString을 활용한 경로 혼잡도 표현에 어려움을 겪음.
- **도로 링크 데이터의 부재** : 10m 단위 도로구간 공간 데이터에 도로 링크 데이터가 존재하지 않아 이를 사용하여 10m 단위의 유동 인구 수를 정확히 파악할 수 없었음. 따라서 전체 도로에 대한 유동 인구를 계산하는 방식으로 모델링 진행.

4. 결론

활용방안 2 - 기존 분석자료를 활용하여 혼잡도 시스템 개선

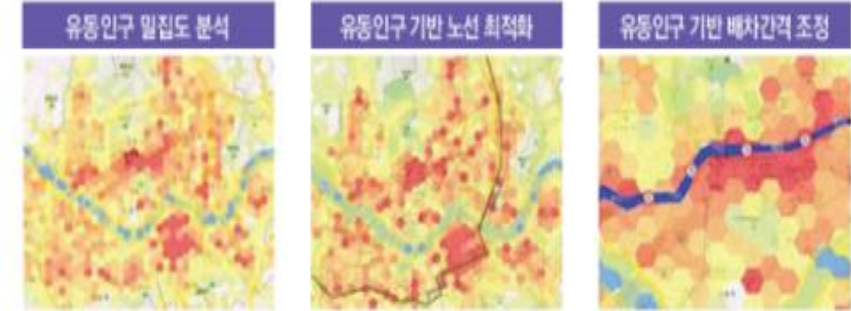
현재 서울시 실시간 도로 데이터의 도로 혼잡도



- 현재 서울시에서 제공하는 실시간 도시데이터는 도로 혼잡도를 권역 단위로만 보여줘, 특정 도로의 실제 유동 인구 수를 정확히 파악하기 어려움.
- 이러한 제한은 10m 도로구간 별로 추정되는 유동 인구 데이터에 대해서도 정확한 값을 추정하는 것이 어렵다는 것을 의미함.

개선 아이디어

[그림1] 빅데이터를 활용한 서울시 심야버스 노선 최적화



[자료: 서울시]

- 최근 도로 교통부에서 빅데이터를 활용하여 서울시 심야버스 노선을 최적화한 사례를 참고해 볼 때, 이와 유사한 방법으로 사람들의 도보 이동경로에 대한 혼잡도를 평가하고 효율적인 경로 계획을 수립하는 것이 가능할 것으로 보임.
- 이를 통해 도로 구간을 더욱 세분화하여 인구 혼잡도를 보다 정밀하게 표현할 수 있음.

사용 데이터 목록

데이터 명		확장자	설명	출처
중심 데이터	서울시 10m 도로구간별 추정 유동인구	Csv	도로구간 별 유동인구 바탕으로 혼잡도 및 변화율 계산	서울시 빅데이터 캠퍼스
지리 및 건물 요인	10m 단위 도로링크	Csv	시각화를 위한 도로 위치 좌표 파악에 사용	서울시 빅데이터 캠퍼스
	서울시 10m 단위 도로구간 공간데이터	Shp		서울시 빅데이터 캠퍼스
	도로경계_면	Shp		국토지리정보원
	전국표준노드링크	Shp		국가교통정보센터
	건물공간정보	Shp	도로 주변 건물 수를 추출하여 파생변수로 사용	서울신용보증재단
교통 요인	국토교통부_전국 버스정류장 위치정보	Csv	도로 주변 버스정류장의 수를 추출하여 파생변수로 사용	국토교통부
기상 요인	조건별 기후 통계분석	Csv	월 평균 기후 추출하여 상관관계 분석을 위한 파생변수로 사용	기상청
축제 사고 요인	행정안전부_통계연보_지역별_지역축제 안전사고	XML	축제 안전사고 발생건수 분석을 위해 사용	행정안전부

분석툴

python™

pandas

matplotlib

Sweetviz

jupyter



MySQL™



Folium

GeoPandas

+ a b | e a u

참고 자료

[한국관광공사] 지역축제 방문객, 팬데믹 이전보다 확 늘었다. 한국관광공사 보도자료, 2023.03.07

김기욱.2022.[다중밀집 인파 사고 예방을 위한 대책 마련 필요] 다중밀집인파사고 예방을 위한 대책 마련 필요. 부산발전포럼,122-128

[서울 정책아카이브] 빅데이터를 이용한 교통계획: 심야버스와 사고줄이기

박예림, 강영옥.2019.통신 데이터를 활용한 도보관광코스 유동인구 추정 및 분석, 한국국토정보공사

윤정미, 최돈정. 2015.서울시 유동인구 분포의 공간 패턴과 토지이용 특성에 관한 지리가중 회귀분석, 대한공간정보학회

채한희, 이경환.2023.상업지역 유동인구에 영향을 미치는 도시공간구조 및 물리적 환경특성 분석, 대한건축학회

김해, 이환필, 권철우, 박성호, 박상민, 윤일수.2018.유동인구 빅데이터 기반 고속도로 휴게소 혼잡지표 개발 연구, 한국ITS학회



감사합니다

2023 서울시 빅데이터 캠퍼스 공모전