



## 2023 빅데이터캠퍼스 공모전 분석결과(요약)

팀명	페스파인더(Festfinder)	접수번호	미기재
제목	스마트 경로 추천 : 우선순위 기반 혼잡도 대응 시스템		
추진배경 및 필요성	<p><b>○ 이태원 참사 이후 인파 사고에 대한 관심 증가</b></p> <p>▶ 이태원 참사는 많은 국민에게 인파사고에 대한 트라우마를 안겼음. 그러나 참사 이후로도 축제기간의 총 방문객 및 일평균 방문객 수는 꾸준히 증가하는 추세. → 또 다른 인파사고의 발생 가능성 우려.</p> <p><b>○ 한번 발생하면 큰 사고로 이어지는 인파 사고</b></p> <p>▶ 인파 사고 : 사람들이 모여드는 장소나 행사에서 발생하는 대규모 충돌이나 진압, 퇴출로 인해 발생하는 대형사고. ▶ 매년 적지 않은 수의 인파 사고가 축제에서 발생하는데, 이러한 사고의 경우 한번 발생 시 작은 원인으로 인해 수많은 사상자가 발생하는 큰 사고로 이어지는 '나비효과'와 같은 현상이 발생.</p> <p><b>○ 기존의 인파사고 방지 대책과 관련 시스템의 보완 및 연구 필요성</b></p> <p>▶ 지능형 인파감지 CCTV : 실제 현장에 가서 혼잡도 정보를 듣지 않는 한, 인구 혼잡도를 사람들에게 제공해 줄 수 없음. ▶ 서울시 실시간 도시데이터 : 큰 틀에서의 인구분포만을 지도로 나타냄. 실제로 사람들이 이용하는 도로의 혼잡 상황을 파악하기는 어려움. ▶ 기존 지도 앱 : 경로 추천 시 최단거리를 우선시하고 실제 인구 혼잡도는 고려하지 않고 같은 경로를 찾는 모든 이용자에게 동일한 경로로 안내하여 도로 혼잡을 유발할 수 있음.</p> <div style="display: flex; justify-content: space-around;">   </div> <p>→ 해당 문제점들을 해결하기 위한 새로운 경로 추천 알고리즘 고안</p>		
1. 투입 데이터	<p>서울시 10m 도로구간별 추정 유동인구 10m 단위 도로링크 서울시 10m 단위 도로구간 공간데이터 (* 서울시 빅데이터 캠퍼스 데이터)</p>		<p>도로경계_면 전국표준노드링크 건물공간정보 국토교통부_전국 버스정류장 위치정보 조건별 기후 통계분석 행정안전부_통계연보_지역별_지역축제 안전사고</p>

## ○ 데이터 정제 및 융합

도로당 유동인구 매칭	파생변수 생성		
출퇴근과 같은 고정적인 혼잡이 아니라 행사와 같은 급격한 인파가 몰리는 혼잡이 위험하므로 <b>주말, 활동시간(8시- 20시)</b> 로 한정 한 뒤 도로당 평균 유동인구로 칼럼 매칭 및 전처리	인구변화율	도로당 유동인구	도로 근처 건물수
	반경 200미터 내 근처 버스정류장 수	기상상태	도로면적

⇒ 유동인구변화를 고려한 혼잡도 계산

데이터 분포 탐색 및 전처리	상관관계 분석

## ○ 데이터 분석 및 모델링

## 유동인구 분석

혼잡도와 시간대 인구변화 관계

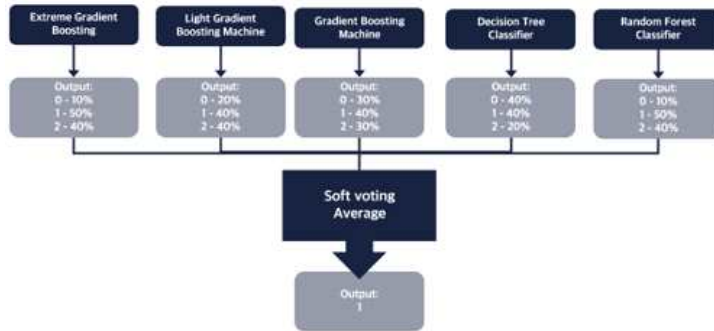
Label	감소	변화	증가	급증
여유	388	191	125	228
보통	14	63	216	146
불편			38	18

혼잡도 분포

월별 유동인구

월별도로의 유동인구를 시각화한 후 유동인구가 많은 지역과 그 지역의 혼잡도를 보여줌

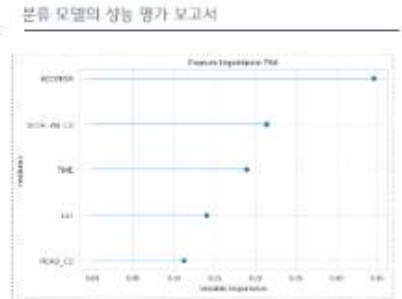
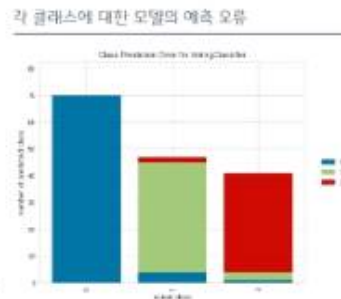
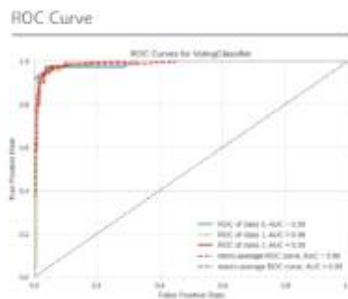
## 유동인구 혼잡도 예측 모델링



Model		Accuracy	AUC	Recall	Prec	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.7578	0.8859	0.7578	0.7609	0.7531	0.6122	0.6184	0.3903
gbost	Extreme Gradient Boosting	0.7668	0.9027	0.7468	0.7531	0.7428	0.5956	0.6001	0.5140
lightgbm	Light Gradient Boosting Machine	0.7668	0.9010	0.7468	0.7454	0.7421	0.5960	0.5998	0.2223
rf	Random Forest Classifier	0.7613	0.8835	0.7413	0.7449	0.7376	0.5862	0.5900	0.4203
dt	Decision Tree Classifier	0.7358	0.8105	0.7358	0.7392	0.7331	0.5819	0.5851	0.1880
knn	K Neighbors Classifier	0.7385	0.8740	0.7385	0.7412	0.7273	0.5763	0.5859	0.2943
gbc	Gradient Boosting Classifier	0.7221	0.8913	0.7221	0.7318	0.7205	0.5582	0.5631	0.7108
ada	Ada Boost Classifier	0.8430	0.7856	0.8430	0.8731	0.8485	0.4495	0.4574	0.2545
qda	Quadratic Discriminant Analysis	0.8295	0.8205	0.8295	0.8522	0.8550	0.4265	0.4310	0.1800
lda	Linear Discriminant Analysis	0.6287	0.8123	0.6287	0.8154	0.8168	0.4040	0.4072	0.2580

- 융합한 데이터들로 유동인구의 혼잡도를 예측할 수 있는지 평가하여 실제로 파생변수들을 통해 혼잡도를 예측하여 추후 혼잡도를 예측하여 경로를 추천하는 시스템 활용으로 나아가고자함
- Boosting 계열 모델을 여러개로 Soft Voting하여 예측력을 높임

## ○ 분석 결과



최종결과,  
활용방안  
및  
기대효과

- ▶ 위치와 시간적 요인이 혼잡도에 상당한 영향.  
→ 월별 데이터 활용 및 지역 범위를 여의도로 한정된 것이 영향을 끼쳤을 것으로 추정
- ▶ 장기간에 걸쳐 축적된 유동인구, 날씨정보, 지역특성 등을 포함한 종합적인 데이터를 활용하여 모델을 학습시킬 경우, 더욱 의미있는 결과를 도출할 수 있을것으로 기대.

## ○ 활용 방안 1- 혼잡도 점수 기반의 경로추천 알고리즘 제안

### ▶ 경로 종합 평가 점수 산식

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

지나가는 도로의 혼잡도를  
합하여 총합을 비교하여  
최적의 경로를 선정

### ▶ 경로 종합 평가 점수 산출 예시



경로	여유	보통	불법	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

## ○ 활용 방안 2- 기존 분석자료를 활용하여 혼잡도 시스템 개선

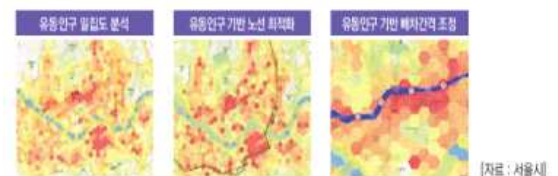
### 기존 서울 실시간 데이터의 도로 혼잡도



- 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 심야버스노선 최적화



- 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 인구혼잡도를 보다 정밀하게 표현할 수 있음.

## ○ 프로젝트 기대 효과

이용자 측면	행정 기관 측면	행사 기획 측면
<ul style="list-style-type: none"> <li>● <b>개인의 안전 향상</b> : 경로 추천을 통해 혼잡한 지역을 회피함으로써 사고 위험, 전염병 노출 위험 감소</li> <li>● <b>시간 절약 및 편의성 증대</b> : 덜 혼잡한 경로 선택함으로써 통행 시간 단축 및 쾌적한 이동 가능</li> </ul>	<ul style="list-style-type: none"> <li>● <b>사고 예방을 위한 사전 조치 가능</b> : 혼잡도 예측 모델을 이용, 사고 발생 가능성 높은 지역 사전 파악 및 선제 조치 가능</li> <li>● <b>교통 혼잡 완화</b> : 혼잡한 경로 회피를 유도하여 교통 흐름 개선 및 혼잡 완화 가능</li> <li>● <b>공공 안전 강화</b> : 혼잡도 정보를 활용해 공공의 안전 유지 및 강화에 기여 가능</li> </ul>	<ul style="list-style-type: none"> <li>● <b>행사 기획 및 배치 최적화</b> : 혼잡도 예측 모델을 활용해 행사 위치 선정, 인원 배치, 물류 계획 등을 효과적으로 수립 가능</li> <li>● <b>안전 사고 예방</b> : 혼잡도 예측 결과를 바탕으로 동선을 확보해 안전사고 예방 가능</li> </ul>

○ **분석툴** : Python, Pandas, Matplotlib, Sweetviz, Jupyter, CSV, MySQL, API, Folium, GeoPandas, Tableau

## ○ 참고문헌

- 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, 강영욱. 2019.통신데이터를 활용한 도보관광코스 유동인구 추정 및 분석, 한국국토정보공사
- 윤정미, 최돈정. 2015.서울시유동인구 분포의 공간 패턴과 토지이용 특성에 관한 지리가중 회귀분석, 대한공간정보학회
- 채한희, 이경환. 2023.상업지역유동인구에 영향을 미치는 도시공간구조 및 물리적 환경특성 분석,대한건축학회
- 김해, 이환필, 권철우, 박성호, 박상민, 윤일수. 2018.유동인구 빅데이터 기반 고속도로 휴게소 혼잡지표 개발 연구, 한국ITS학회

분석툴,  
참고문헌