

# 감기 예방을 위한 길 찾기 서비스

삼인조

# 목차

1. 서비스의 필요성
2. 서비스 개요
3. 사용한 데이터
4. 데이터 전처리
5. 발생건수 예측 모델 개발
6. 웹페이지 개발
7. 보완점 및 개선점

# 서비스의 필요성

- ✓ 엔데믹 이후 감기 등 전염병 발생건수 다시 증가하는 추세
- ✓ 감염병 확산에 대한 대응이 아닌 예방이 필요
- ✓ 대중교통 이용 시 밀폐된 공간에 장시간 머무르게 되며, 이는 특히 호흡기성 감염병에 노출될 가능성을 높임
- ✓ 개인적인 차원에서 혼잡한 시간대와 장소를 피할 수 있는 수단을 마련하고자 함
- ✓ 기존의 길 찾기 서비스와 열차 혼잡도를 결합한 서비스 개발

# 서비스 개요

- 히트맵과 도표를 이용해 당일 서울시 행정구별 감기 예측건수 시각화
- 사용자의 이동 경로를 제공함과 동시에  
지하철을 이용하는 경로에서는 지하철 혼잡도 정보 함께 제공
- 사용자에게 출발지와 도착지의 감기 예측건수와 지하철 혼잡도 정보를  
제공
- 상대적으로 안전한 경로 선택의 기회 제공

# 사용한 데이터

- 국민건강보험공단\_진료건수 정보
- 관심도가 높은 4대 질병(눈병, 천식, 감기, 피부염) 시군구별 진료건수 제공
- 2014년~2023년 9월 데이터 취합
- 서울시의 감기 진료건수 데이터 추출하여 사용

## CSV 국민건강보험공단\_진료건수 정보

[↓ 다운로드](#)

국민건강보험공단 진료건수 정보

-수요조사 결과에 따라 관심도가 높은 4대 질병(눈병, 천식, 감기, 피부염) 시군구별 과거 진료건수 제공

\* 2023.5.1. ~ 2023.9.30. 자료파일, 2023.5.1.이전 자료는 과거 파일 참고

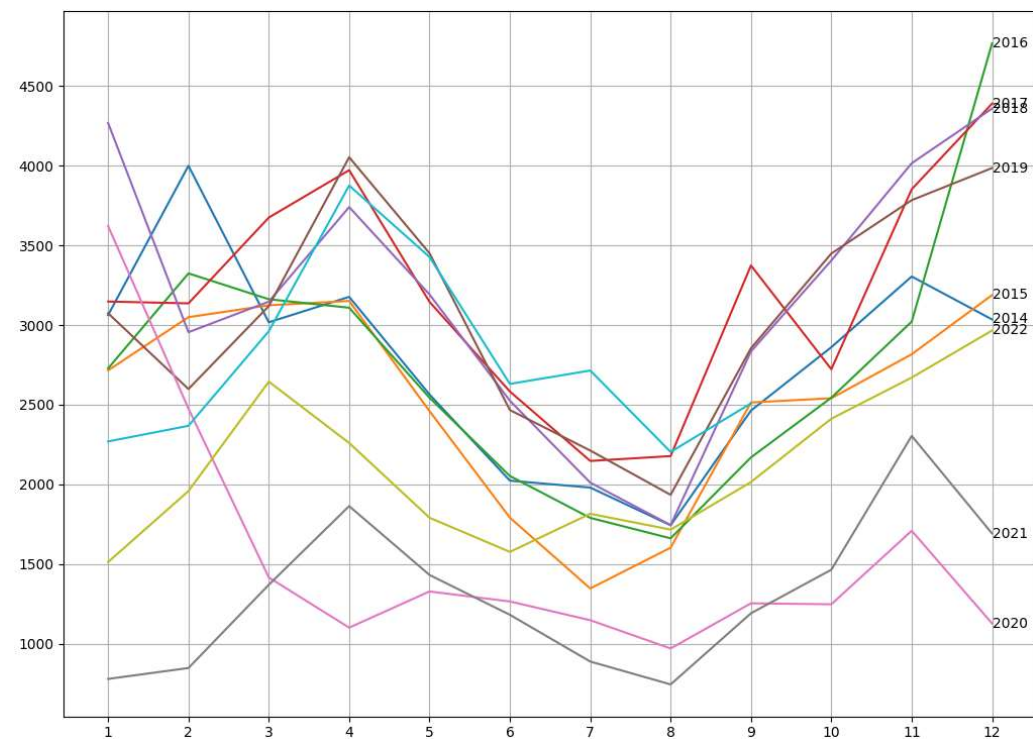
\* 국민건강알람서비스 종료로 인해 2024.2.29.까지 데이터 제공

👍 0

👎 0

🔖 관심

# 서울시 연도별 월별 감기 발생건수 분석



- 2014년~2023년의 월별 평균 감기 발생건수 분석

- ✓ 4월과 12월에 가장 많은 발생 건수
- ✓ 8월에 가장 적은 발생 건수
- ✓ 감기 발생건수 변동 패턴을 반영하기 위해 발생건수 양상이 다른 연도는 학습 데이터셋에서 제외 (2014년, 2020년, 2021년, 2022년)

# 감기 발생건수 예측 모델 개발

- 앙상블 모델
- 다양한 패턴 반영, 예측 성능 향상, 과적합 방지 등의 이유로 선택
- RandomForest, GradientBoost, XGBoost 3개 모델의 앙상블
- 시계열성 데이터의 특성을 반영하기 위해 차분 기법 활용
- 행정구별 예측 그래프와 성능 평가 지표 추출
- 주요 평가 지표:  $R^2$

# 앙상블 모델 - 행정구별 예측

```
# 구별로 학습 및 테스트
for district_name, district_code in district_code_map.items():
    # print(f"\n### {district_name} ###")

    # 해당 구의 데이터만 필터링
    district_data = data[data['county_code'] == district_code].copy() # .copy()로 데이터 복사

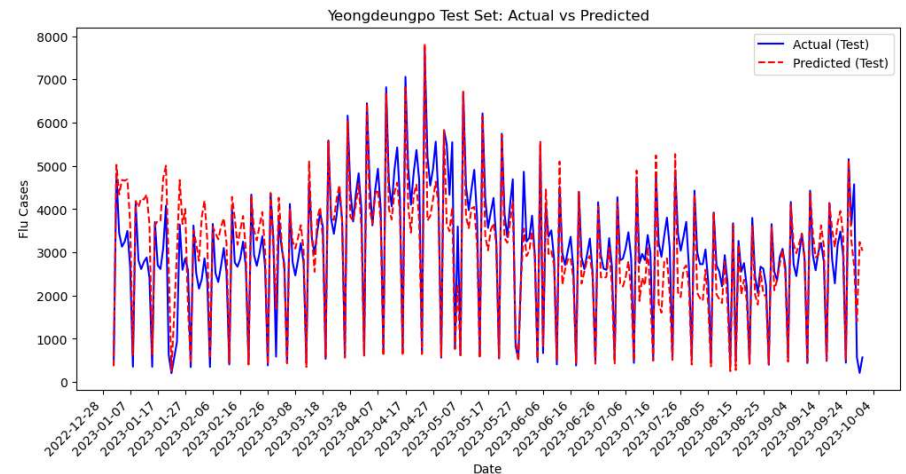
    # 구별 데이터에서 차분 기법 적용 (각 구에 대해 차분 계산)
    district_data['diff_cold_cases'] = district_data['number of cold case'].diff().fillna(0)

    # 데이터셋 분리
    train_df = district_data[(district_data['year'] >= 2015) & (district_data['year'] <= 2018)] # 2015~2018년 데이터로 학습
    val_df = district_data[(district_data['year'] == 2019)] # 2019년 데이터를 검증에 사용
    test_df = district_data[(district_data['year'] == 2023)] # 2023년 데이터를 테스트에 사용

    # 훈련, 검증, 테스트 데이터 준비
    X_train = train_df[['year', 'month', 'day', 'weekday', 'diff_cold_cases']]
    y_train = train_df['number of cold case']
    X_val = val_df[['year', 'month', 'day', 'weekday', 'diff_cold_cases']]
    y_val = val_df['number of cold case']
    X_test = test_df[['year', 'month', 'day', 'weekday', 'diff_cold_cases']]
    y_test = test_df['number of cold case']

    # 모델 정의 및 학습
    best_rf = RandomForestRegressor(n_estimators=100, max_depth=20, min_samples_split=10, min_samples_leaf=2, random_state=42)
    best_gb = GradientBoostingRegressor(n_estimators=200, learning_rate=0.1, max_depth=7, random_state=42)
    best_xgb = xgb.XGBRegressor(n_estimators=100, learning_rate=0.1, max_depth=7, reg_alpha=0.01, reg_lambda=0.01, random_state=42)

    stacking_model = StackingRegressor(estimators=[('rf', best_rf), ('gb', best_gb), ('xgb', best_xgb)], final_estimator=Ridge(), cv=5)
    stacking_model.fit(X_train, y_train)
```



- ✓ 행정구별로 데이터 분리하여 학습
- ✓ 앙상블 구성 모델별 최적의 파라미터 탐색하여 모델에 적용
- ✓ 데이터셋 분리
  - 훈련 데이터: 2015년~2018년
  - 검증 데이터: 2019년
  - 테스트 데이터: 2023년 1월~9월

영등포구의 테스트 데이터 예측 결과

Test MAE: 483.11

Test RMSE: 679.27

Test MAPE: 26.67%

Test  $R^2$  : 0.80



# 웹페이지 개발

- 모델을 이용해 예측한 감기 발생건수를 csv 파일로 저장
- 발생건수 히트맵, 도표와 연동해 시각화
- 경로 탐색에는 T맵에서 제공하는 무료 api 활용
  - TMAP API: Geocoding (주소 → 좌표 변환)
  - TMAP 대중교통 API: 대중교통(경로 탐색 정보), 진입 역 기준 열차 혼잡도
- 출발지/도착지의 예측 감기 발생건수와 열차 혼잡도 정보 확인 가능

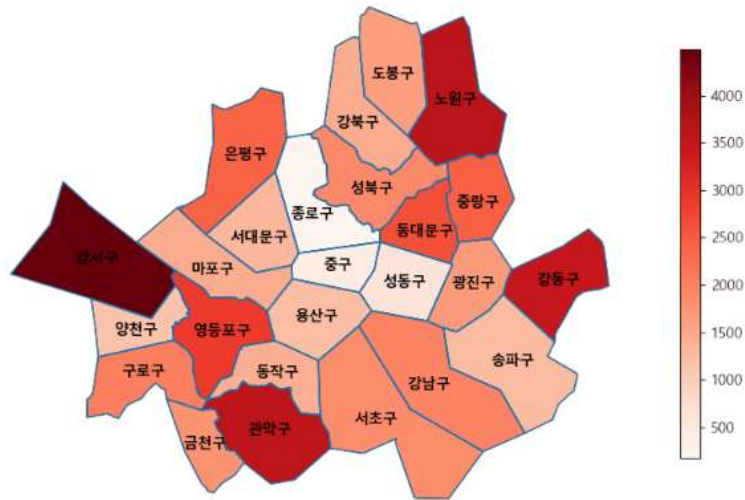
# 웹페이지 개발



서울시 지도를 조회하고 출발지/도착지 주소를 입력

# 웹페이지 개발

## 지도 조회



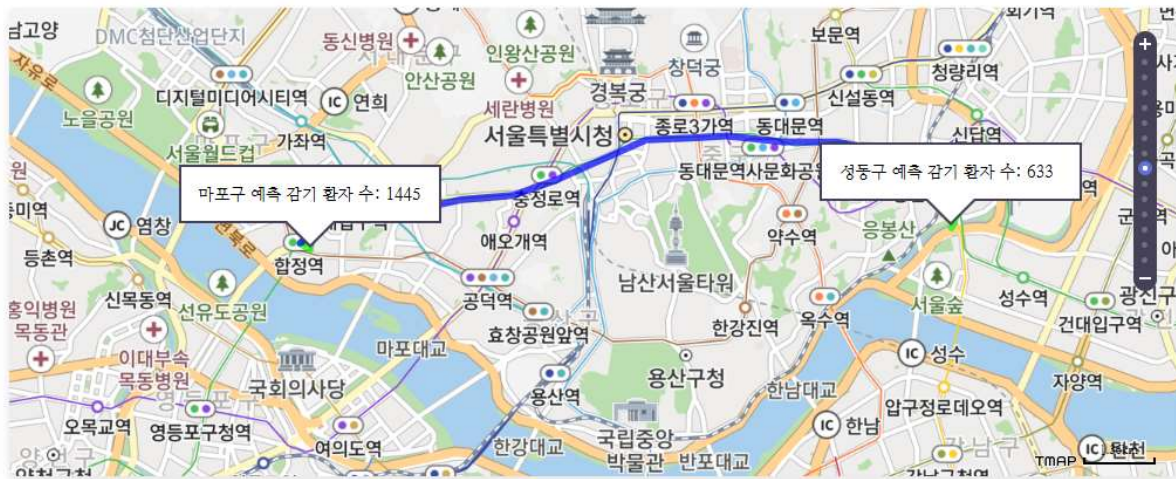
- ✓ 예측한 당일의 감기 발생 건수를 히트맵으로 표현
- ✓ 감염 위험 정도를 직관적, 시각적으로 파악 가능
- ✓ 발생 건수 값을 나타낸 도표도 함께 확인할 수 있음

## 확진자 예측

구역	확진자 수	구역	확진자 수
강남구	1988	강동구	3499
강북구	1453	강서구	4494
관악구	3602	광진구	1680
구로구	2041	금천구	1773
노원구	3647	도봉구	1635
동대문구	2651	동작구	1422
마포구	1445	서대문구	1318
서초구	1854	성동구	633
성북구	1925	송파구	1249
양천구	1128	영등포구	2885
용산구	1218	은평구	2405
종로구	178	중구	419
중랑구	2438		

# 웹페이지 개발

## • 출발지/도착지 검색 후 화면



### 경로 1

총 시간: 42분, 요금: 1500원

타는 역: 합정 - 혼잡도: 혼잡

내리는 역: 한양대 - 혼잡도: 매우 혼잡

### 경로 2

- ✓ 다양한 경로 중 선택 가능
- ✓ 경로 선택 시 확인할 수 있는 사항
  - 이동 소요 시간
  - 요금
  - 출발지와 도착지의 예상 감기 발생건수
  - 타는 역과 내리는 역의 혼잡도
  - 혼잡도 구간: 매우 혼잡, 혼잡, 보통, 여유 4구간으로 분할

# 보완점 및 개선점

- 서울시를 포함한 전국적인 서비스 제공
- 경로 상 모든 지하철역의 혼잡도 정보 제공
- 감기 발생 건수 예측 시 더 다양한 독립 변수 활용 – 예측 성능 증가
- 감기뿐만 아니라 다양한 감염성 질병의 예측건수 제공
- 지하철에 국한되는 것이 아닌 다양한 대중교통 수단에서의 혼잡도 정보 제공