

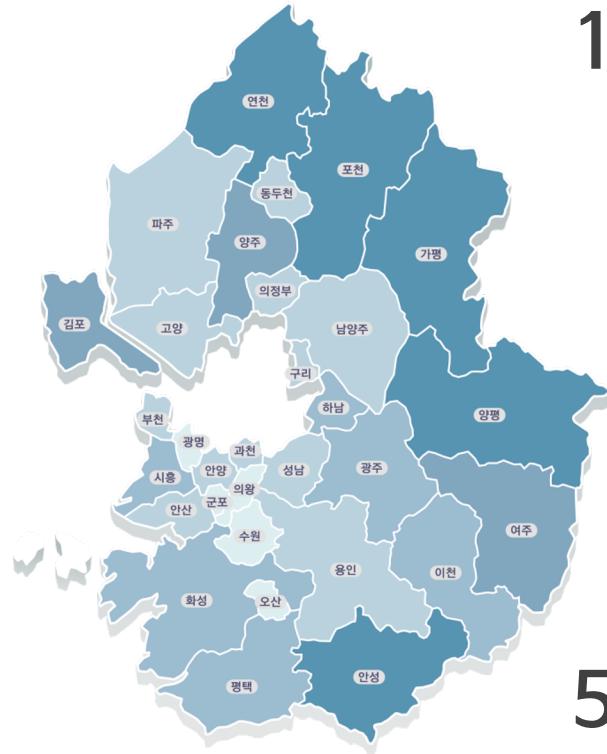
[데이터 기반 코로나19 예측 공모전]

머신러닝을 통한 경기도 확진자수 예측 및 정책 제언

[더조은멤버들]
김우현 . 배정언
안은지 . 엄희진

새로운 경기 ▶ 공정한 세상

목차



1. 요약

2. 가설 및 변수 설정

3. 활용데이터 및 가공

4. 모델구현

5. 정책제언

1. 요약

새로운 경기 ▶ 공정한 세상

참여 개요

팀명 및 대표자 성명	더조은멤버들 / 김우현		참여 구분	팀
팀 구성원 성명	배정언, 안은지, 엄희진		소속 (학교/연구소/기업명)	더조은 컴퓨터 아카데미
대표자 전화번호	010-7565-9519		대표자 email 주소	inco123@naver.com
모델링 정보 요약	개발 환경	파이썬(3.6.10) 아나콘다(4.8.2) R(4.0.2) 주피터노트북(6.0.3) MySQL(5.7)		
	활용 데이터	상세내용 APPENDIX 참고(page16)		
	모델 소개	1. Tensorflow keras LSTM 모델 input : 검사자 수, 정책 변수, 2차공격률 / output : 확진자 수 2. Facebook Prophet 모델 input: 확진일, 누적 확진자 수 / output : 확진자 수		

환자 예측 결과

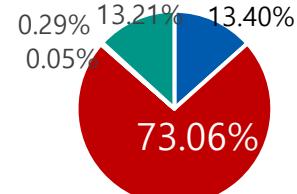
일자	9/30	10/1	10/2	10/3	10/4
환자 수	47	47	46	47	41

2. 가설 및 변수 설정

새로운 경기 ▶ 공정한 세상

경기도 확진자 데이터 분석 및 가설 설정

[경기도 확진자 감염경로]



접촉에 따른 가설 수립 및 국내 감염 확진자로 검증 기준 설정

[가설 설정]

접촉을 줄일수록 확진자 수는 감소할 것이다.

(*확진자 연번 1~4194 대상)

국내 감염 확진자

[국내 확진자 감염 관련성]

(*확진자 연번 1~4194 대상)



- 국내감염자의 경우 확진자 접촉이 가장 높은 수치기록
- 다양한 집단 모임에서 감염 발생

▶ 국내 감염의 경우 접촉이 확진과 연관이 높다고 가정
접촉 관련 변수 선정 후 확진자 수와 관련성 검증

해외 감염 확진자

[입국일 - 확진일간 기간에 따른 확진자 수]

(*확진자 연번 1~4194 대상)



- 해외입국 후 감염된 확진자의 97%가 자가격리 기간(14일) 내 확진 판정
- 타인에게 전파할 위험성 낮음, 거리두기 정책 등 접촉 요소에 대한 영향력 낮음

▶ 해외 확진자의 경우 접촉과 연관성이 낮아 **가설 검증 제외**
해외 확진자 수 추이 기반으로 별도 시계열 분석

2. 가설 및 변수 설정

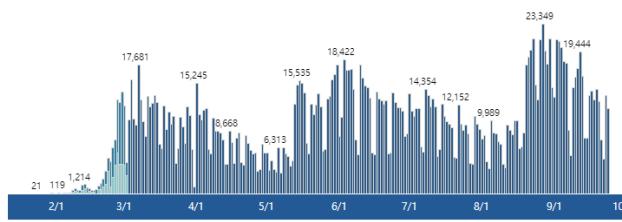
새로운 경기 ▶ 공정한 세상

변수 선정 기준

데이터 유무 + 정부발표 + 기사를 통해 접촉 관련 변수 선정 및 수치화

I. 일별 검사자

확진자와 접촉有 = 검사대상자



검사자 수는 확진자와의 접촉이 의심되어
검사를 받았다고 가정
일별 전국 검사자 수 변수 추가

일별전국검사자수=접촉자수

II. 사회적 거리두기 정책

거리두기정책단계에 따라 가중치 부여

방역수칙 단계별 조치 내용	사회적 거리두기		
	1단계 (생활속거리두기)	2단계	3단계
집합·모임·행사	8종 (방역수칙 준수 강화)	실내 50인·실외 100인 이상 금지	10인 이상 금지
스포츠 행사	참석 관중 제한	무관중 경기	경기 차지
공공 다중시설	운영 하용 (필요시 일정 충단·제한)	운영 중단	운영 중단
민간 다중시설	운영 하용 단, 고위험시설 운영 자체 명령(방역수칙 준수)	고위험시설 운영 중단, 그 외 시설 방역수칙 준수 강제화	고·중위험시설 운영 중단, 그 외 시설 방역수칙 준수 강제화
학교·유치원·어린이집	등교·원격 수업	등교 원격수업 (등교 인원 축소)	원격 수업 또는 휴업
공공 기관 및 기업	유연·재택근무 등을 통한 근무밀집도 최소화 (전 인원의 1/3)	유연·재택근무 등을 통해 근무 모범 제한 (전 인원의 1/2)	필수인원 외 전원 재택근무
민간 기관 및 기업	유연·재택근무 등 통해 균형화 조정	유연·재택근무 등을 통해 균형화 조정	필수인원 외 전원 재택근무 권고

* 다중이용시설 운영 제한 범위는 상황에 따라 탄력적으로 조정 가능 연합뉴스

[이미지 출처]
연합뉴스

- ~ 3/21 : 사회적 거리두기 없음 => 0
- 3/22 ~ 4/19 : 강력한 거리두기 시행 => 1.5
- 4/20 ~ 8/18 : 완화한 형태 거리두기 시행 => 1
- 8/19 ~ 8/29 : 수도권 거리두기 2단계 => 2
- 8/30 ~ 9/13 : 사회적 거리두기 2.5단계 => 2.5
- 9/14 ~ : 사회적 거리두기 2단계 => 2

III. 2차 공격률

유/무증상에 따른 2차 공격률 계산

동아사이언스

방역당국이 첫 언급 코로나19 '2차 공격률'이란..."환자 한명
이 2차 감염 유발하는 수치"

기사입력 2020.06.10 오후 6:51 | 기사원문 | 스크랩 | 봉은글기 · 설정

권준욱 중앙방역대책본부 부본부장은 9일 충북 오송 질병관리본부에서 열린 정례브리핑
에서 "무증상" 경우에 밀접접촉자의 2차 공격률이 약 0.8%밖에 안 된다"며 "영증일 때 2
차 공격률이 3.5%, 증상이 조금 심해지면서 이 비율이 5.7%까지 올라간다"고 말했다.

- 무증상공격률 : 0.8% → 당일 확진자 유/무증상으로
가중치를 부여하여 수치화
- 경증환자공격률 : 3.5%
- 중증환자공격률 : 5.7% → 증상 악화 전 격리 치료를
받을 가능성이 크다고 판단,
중증환자의 공격률은 제외

$$\text{2차공격률} = (\text{무증상} * 0.8\%) + (\text{유증상} * 3.5\%)$$

2. 가설 및 변수 설정

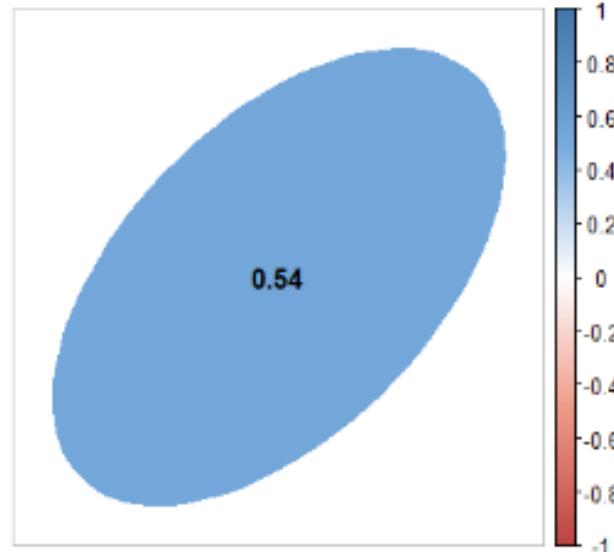
새로운 경기 ▶ 공정한 세상

변수 적합성 검증

‘일별 검사자 수, 사회적 거리두기 정책, 2차 공격률’은 확진자 수와 연관성 입증됨

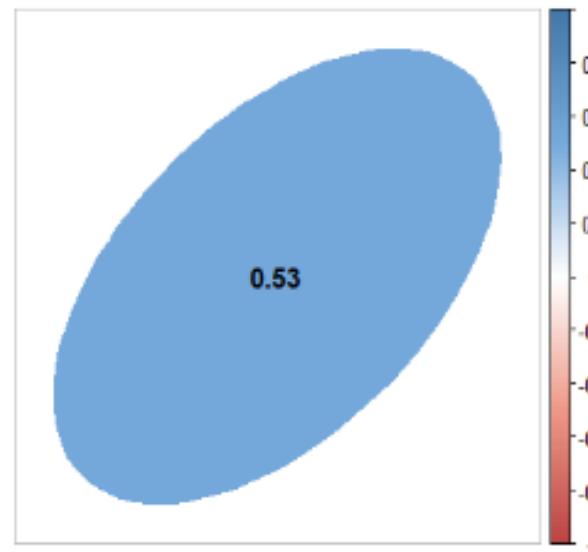
〈Pearson 상관분석을 통해 ‘확진자 수’와 변수간 상관관계 파악〉

I. 확진자 수와 일별 검사자 수의 관계



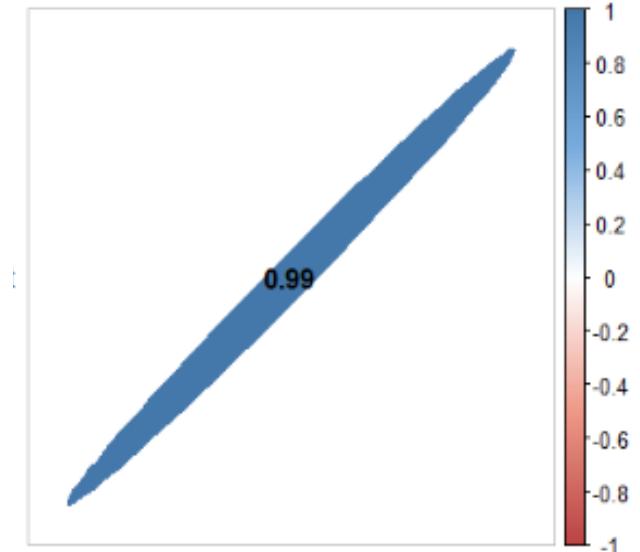
다소 높은 상관관계

II. 확진자 수와 사회적거리두기 정책의 관계



다소 높은 상관관계

III. 확진자 수와 2차 공격률의 관계



매우 높은 상관관계

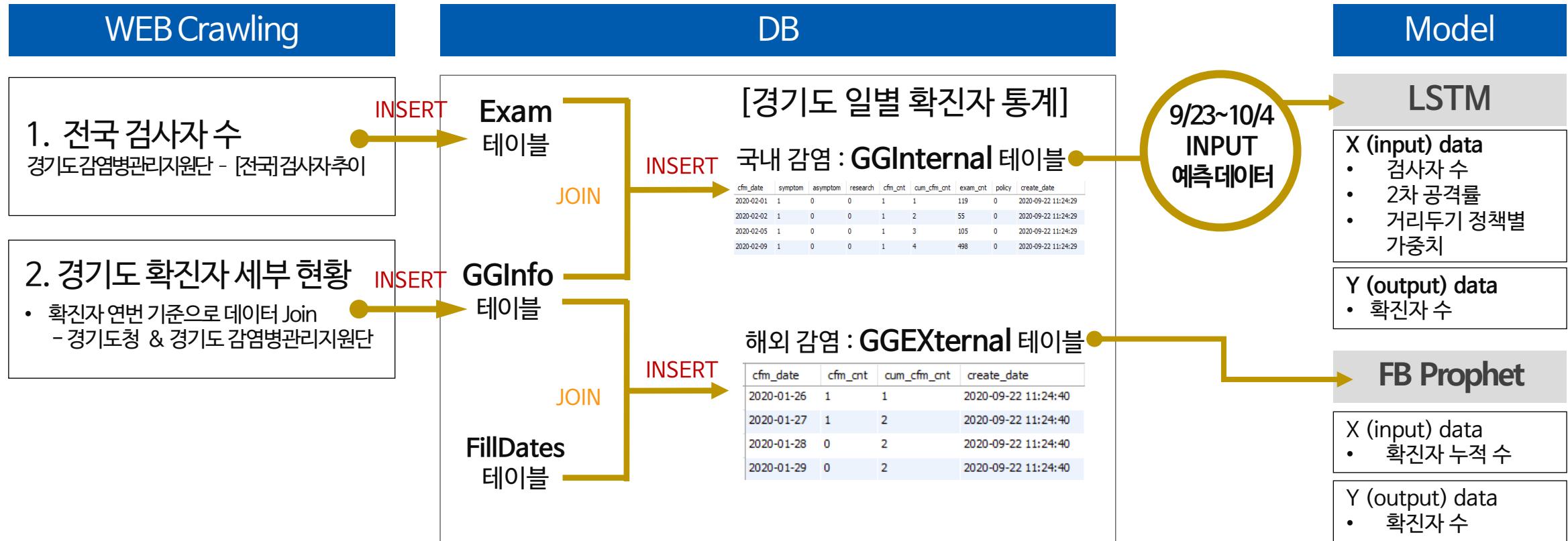
위의 결과를 통해, ‘접촉을 줄일수록 확진자 수는 줄어든다’는 가설 입증

3. 활용 데이터 및 가공

새로운 경기 ▶ 공정한 세상

전체 데이터 가공 프로세스

웹 크롤링을 이용한 DB 적재 자동화 시스템 구축



3. 활용 데이터 및 가공

새로운 경기 ▶ 공정한 세상

데이터 수집 및 적재

1/26 ~ 9/22 까지 경기도 확진자 데이터 수집 및 DB 적재

웹 크롤링

1. Selenium을 이용한 데이터 크롤링

4224건										
번호	전국번호	확진일자	지역(선택)	발생경위	관련성	등록일	수정일	삭제여부	등록자	수정자
4224	화인중	9.22	안양시	9.21일 오현 및 복동 등 증상	미분류	2020. 9. 23.	2020. 9. 23.	N	김민경	김민경
4223	화인중	9.22	수원시	확진자 접촉, 무증상	확진자와 접촉자	2020. 9. 23.	2020. 9. 23.	N	김민경	김민경
4222	화인중	9.22	의왕부시	확진자 접촉, 9.17일 근처동 등 증상	신촌 세브란스/속초	2020. 9. 23.	2020. 9. 23.	N	김민경	김민경

3. 코로나19 경기도 확진자 (상세)

번호	확진자	성별	나이(세)	연령대	확진일자	증상발현일	무증상/조사중	경기번호	지역	제작일	감염경로
1	3	M	54	50	2020. 1. 26.	2020. 1. 22.	고양1	고양발신서	해외유입		
2	4	M	55	50	2020. 1. 27.	2020. 1. 25.	평택1	평택	해외유입		
3	12	M	48	40	2020. 1. 1.	2020. 1. 20.	부천1	부천	해외유입		

2. BeautifulSoup 이용하여 크롤링, 데이터 전처리 튜플에 담아 List에 append

```
data = (cfmDate, row[2])
exam_list.append(data)
```

```
[('2020-09-23', '11677'), ('2020-09-22', '13462'), ('2020-09-21', '4818'), ('2020-09-20', '7457'), ('2020-09-19', '12687'), ('2020-09-18', '14347'),
```

3. List를 DB로 INSERT

```
sql_exam = "insert ignore into Exam(exam_date, exam_cnt) values(%s,%s)"
cur.executemany(sql_exam, exam_list)
```

실행 :: gg_main.py

DB

1. 국내/해외 구분

- GGInfo 테이블 - how_inf(상세확진자내용)에서 <유럽/해외/입국/비행기/태국/이란> 키워드 검출 후 국내외 분류

```
-- 국내
select * from gginfo where gginfo_id not in (
select gginfo_id from GGInfo where how_inf REGEXP '유럽|해외|입국|비행기|태국|이란' or relation REGEXP '유럽|해외|입국|비행기|태국');

-- 해외
select * from GGInfo where how_inf REGEXP '유럽|해외|입국|비행기|태국|이란' or relation REGEXP '유럽|해외|입국|비행기|태국';
```

2. 국내 감염 : GGInternal 테이블

- GGInfo 테이블 : 2/1 ~ 9/22 까지 확진자 발생일 기준으로 데이터 기입
- GGInfo 테이블의 issymptom 컬럼에서 유증상/무증상/조사중 구분
- 당일 확진자 수 & 누적 확진자 수 컬럼 계산
- GGInfo / Exam 테이블을 확진일 기준으로 LEFT JOIN
- 기간에 맞게 정책 컬럼 UPDATE

3. 해외 감염 : GGExternal 테이블

- FillDates 테이블 : 1/26 ~ 9/22까지의 연속 날짜 기입 (확진자 0인 날짜 포함)
- FillDates / GGInfo 테이블을 확진일 기준으로 LEFT JOIN
- 당일 확진자 수 & 누적 확진자 수 컬럼 계산

※ Null 값은 0으로 처리

실행 :: insert_table.sql

3. 활용 데이터 및 가공

새로운 경기 ▶ 공정한 세상

INPUT데이터 예측

9/23 ~ 10/4기간 확진자 수 예측을 위한 INPUT데이터 예측

1. 예측 모델 선정

[Numpy / 엑셀 /FB Prophet 이용 9/13 ~ 19일 검사자수 예측]

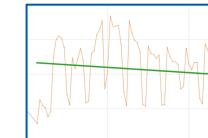
기간	실제 검사자수	np.polyfit()	EXCEL	FB
2020.9.13	7,692	14,170	15,022	8,472
2020.9.14	7,623	13,959	14,897	9,477
2020.9.15	13,470	13,748	14,772	9,860
2020.9.16	14,141	13,537	14,646	11,392
2020.9.17	12,907	13,326	14,521	11,009
2020.9.18	14,347	13,114	14,396	10,626
2020.9.19	12,687	12,903	14,270	10,243

예측 도구 중 가장 정확도가 높은
Python np.polyfit() 설정

2. 추세선 분석



신천지 (80일)
기울기: -42
절편: 9931



이태원 (100일)
기울기: -25
절편: 11651



사랑제일교회&광화문 집회
기울기: 116
절편: 10096

3. 미래 INPUT데이터



사랑제일교회&광화문 집회
기울기, 절편으로
INPUT데이터 예측

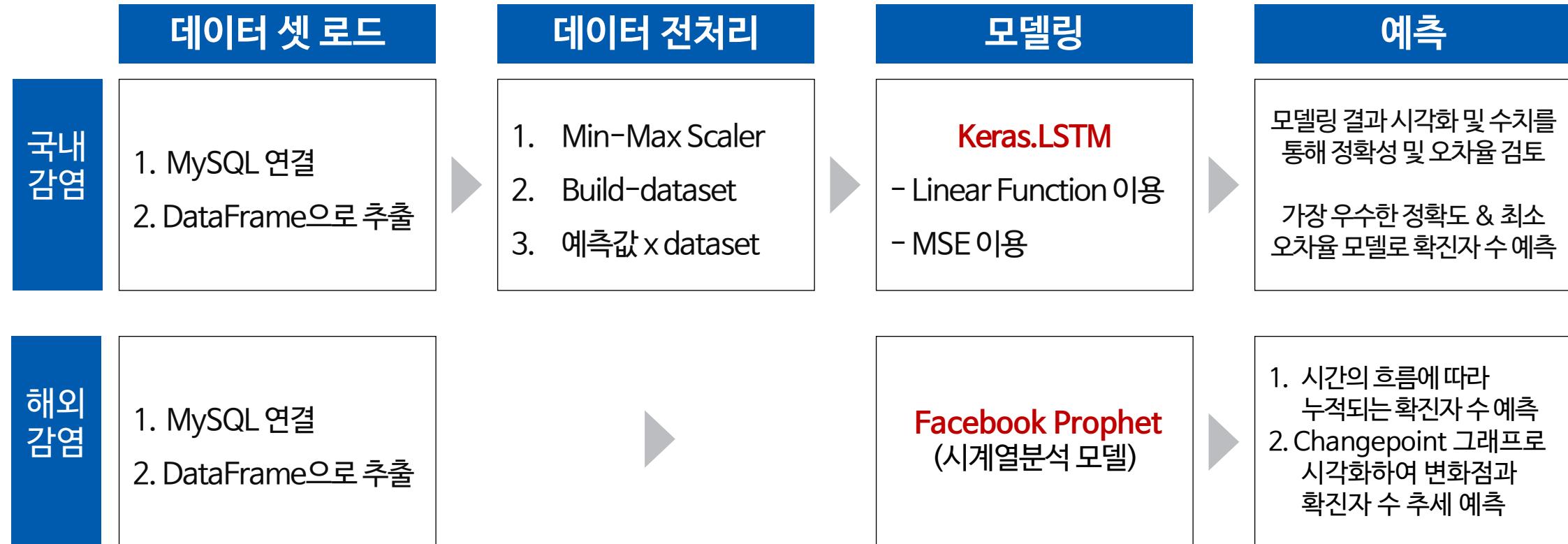
두 차례 대유행 구간이 비슷한 추이를 보임
현재 사랑제일교회&광화문 집회 대유행도
동일 흐름을 보일 것으로 추측

4. 모델 구현

새로운 경기 ▶ 공정한 세상

모델 구현 프로세스

머신러닝을 이용한 경기도 확진자 수 예측 프로세스 구축



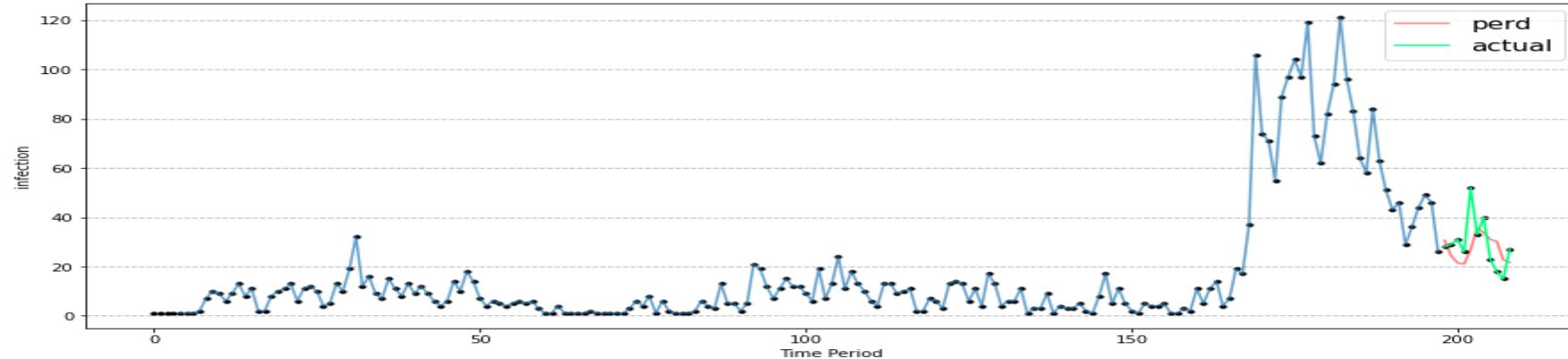
4. 모델 구현

새로운 경기 ▶ 공정한 세상

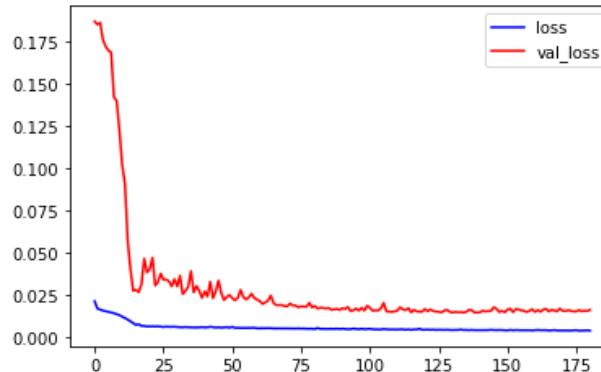
경기도 국내 감염 확진자 수 예측 모델

LSTM 이용한 시계열 예측

〈전체 구간〉 2/1~9/22까지 확진자 수 데이터 흐름 및 모델링 완성도 파악

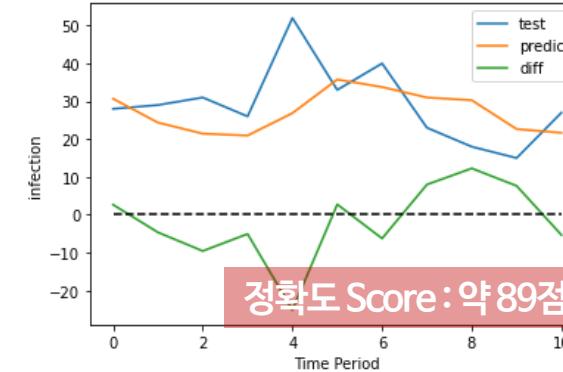


〈Train vs Test Loss-rate〉



학습 데이터셋 VS 테스트 데이터셋
최소 오차율 최적화

〈실제값 vs 예측값 비교〉



테스트 데이터 vs 실제 데이터
차이로 정확도 파악

〈예측 결과〉

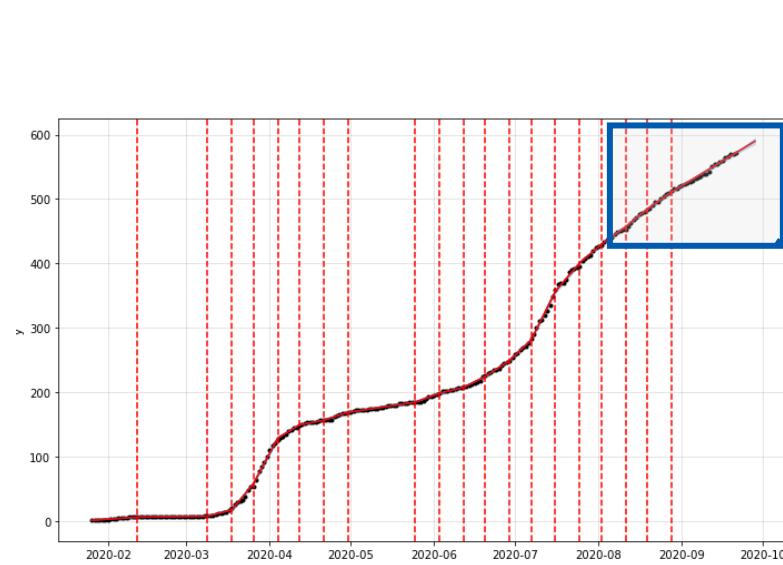


가장 높은 정확도 점수 모델에
예측 데이터셋을 적용, 예측의 결과 확인

4. 모델 구현

새로운 경기 ▶ 공정한 세상

경기도 해외유입 확진자 수 예측 모델



Facebook Prophet 이용한 시계열 예측



Dataset :: GGExternal 테이블

Sorce :: ggexternal_prediction.ipynb

경기도 해외유입 확진자 수 최종 결과

일자	9/30	10/1	10/2	10/3	10/4
국내	45	44	44	44	39
해외	2	3	2	3	2
총합	47	47	46	47	41

5. 정책 제언

새로운 경기 ▶ 공정한 세상

정책 배경

사회적 거리두기 정책 강화로 인한 시민, 상인의 불만 증가 시민, 상인, 정부 모두 WIN-WIN 할 수 있는 정책 필요

- 방역 조건을 높일수록 영업 제한 발생
- 경제적 손실로 인한 불만 및 답답함 가중

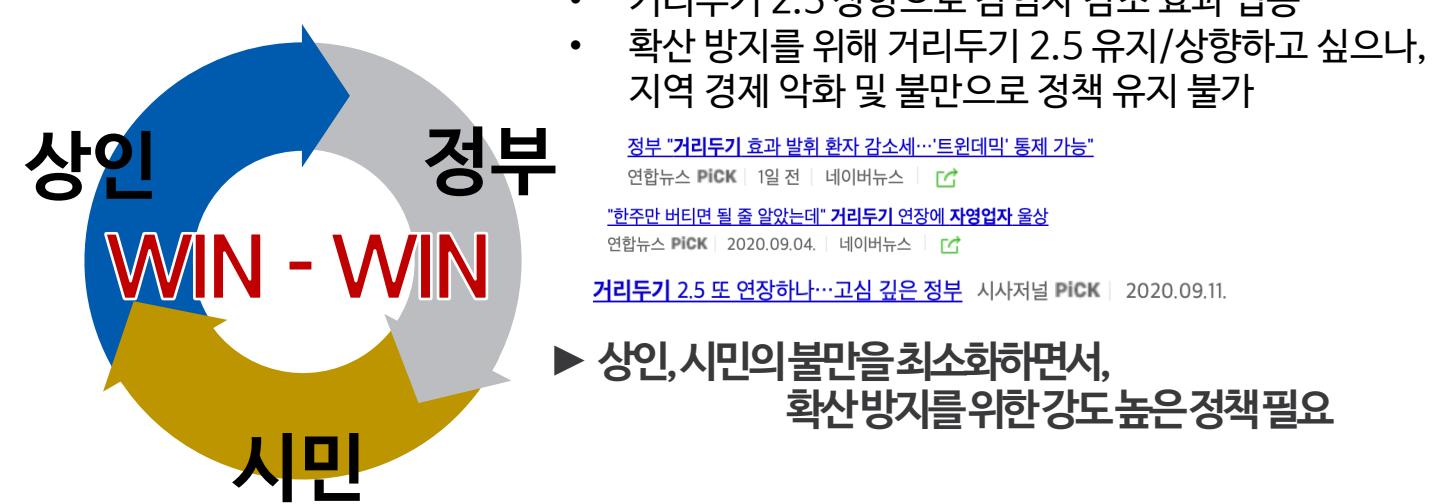
[거리두기 2.5단계 2주째...텅텅 빈 식당·폐업하는 상점](#)
YTN PICK | 2020.09.11. | 네이버뉴스 |

[\[광주24시\] 광주 '거리두기 준3단계' 연장에 자영업자 집단 반발](#)
시사저널 | 2020.09.11. | 네이버뉴스 |

["거리두기 2.5단계...수도권 소상공인 매출 급감 직격탄"](#) 브레이크뉴스

사회적 거리두기 강화...수원 학원가는 불만, 변화가는 수심 가득
뉴스1 | 2020.08.31. | 네이버뉴스 |

▶ 경제적 활동이 보장된 방역 정책 필요



[정부 '거리두기' 효과 발휘 환자 감소세...·'트원데믹' 통제 가능](#)
연합뉴스 PICK | 1일 전 | 네이버뉴스 |

["한주만 버티면 될 줄 알았는데" 거리두기 연장에 자영업자 올상](#)
연합뉴스 PICK | 2020.09.04. | 네이버뉴스 |

[거리두기 2.5 또 연장하나...고심 깊은 정부](#) 시사저널 PICK | 2020.09.11.

▶ 상인, 시민의 불만을 최소화하면서, 확산 방지를 위한 강도 높은 정책 필요

- 시설 이용 제한의 불편함 가중 및 역으로 야외로 이동하여 밀집 형성
- 실내 활동 장기화로 코로나 블루 발생 및 감염에 대한 두려움으로 야외 활동 기피

▶ 무조건적인 통제보다는 시민의 자율성이 보장된 방역 정책 필요

[연트럴 단속하자 '책거리'로...길맥족 자정까지 '야외 출판'](#)
뉴스1 PICK | 2020.09.11. | 네이버뉴스 |

[실내 취식 금지되자 한강공원 '복새통'...거리두기 무색](#)
SBS PICK | 2020.09.05. | 네이버뉴스 |

[술집·식당 문 닫으니 야외 출판...방역 풍설효과 뻗간불](#)
국민일보 | 2020.09.11. | 네이버뉴스 |

['코로나 블루' 심각...20대 우울증 환자, 4년 새 두 배 증가](#)
한국경제 | 1시간 전 | 네이버뉴스 |

5. 정책 제언

새로운 경기 ▶ 공정한 세상

자영업자 대상 정책 제안

〈코로나 안심 존〉 인증 발급. 인증 조건에 만족한 상점에 한해 오후 9시 이후 영업 허가 영업 연장을 원하는 자영업자의 자발적인 방역 유도

[코로나 안심 존 인증 조건 예]

1. 책임자 또는 종사자는 마스크 필수 착용 및 개인 위생 준수
2. 고객 입장 시 마스크 착용 및 발열 체크
(음식, 음료 섭취 외에 마스크 착용 안내)
3. 손 소독제 구비 및 에티켓 준수 안내문 부착
4. 고객 출입 명부 작성 및 관리
5. 테이블이 구비된 상점(예: 음식점, 카페, PC방)의 경우
테이블 간막이 설치 또는 테이블 거리 두기
6. 대형 상가의 경우 경기도에서 지원하는 전문 방역
소독 실행(주 1회)
7. 지속적으로 접촉이 발생하는 부위
(손잡이, 스위치, 변기 버튼 등) 일 2회 소독
8. 대기자 발생 시 최소 1m 이상 간격 두기 안내

[인증 마크 발급 예]



- 조건에 만족하는 상점 대상 인증 발급
- 인증 마크 상점 입구 부착
- 오후 9시 이후 영업 허가
- 인증 마크 없이 9시 이후 영업 시 벌금 청구
- 매주 단속을 통해 방역 위반 시 인증 철회

[기대효과]



- 강압적인 영업 제한이 아닌 방역 준수에 따른 영업 허가로 상인 불만 감소
- 인증 마크를 받기 위해 상인 스스로 자발적으로 방역에 힘씀. 감염 확산 방지
- 인증 마크를 확인한 고객 역시 안심하고 상점 이용 가능 및 신뢰 형성으로 이용 빈도 증가 예상

5. 정책 제언

새로운 경기 ▶ 공정한 세상

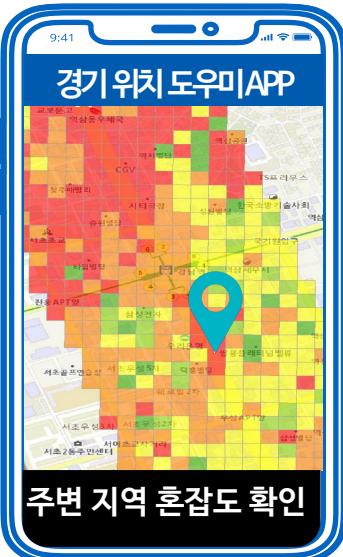
시민 대상 정책 제안

실시간 유동인구 파악 가능한 서비스APP 제공 인구 밀집 지역 방문을 자제하도록 시민들의 자발적인 거리두기 행동 유도

[시행 가능 서비스 제안]

1. 실시간 밀집도 확인

서비스 앱을 통해 특정 지역 밀집 확인 가능
=> 경각심을 유도하여 방문을 자제하도록 함



[이미지출처] 지오비전 소개서

2. 재난 알림 발송

앱을 설치하지 못하는 시민들을 위해
밀집도에 따라 **여유** / **보통** / **혼잡** 3단계 분류
=> **혼잡** 지역 진입 시 경고



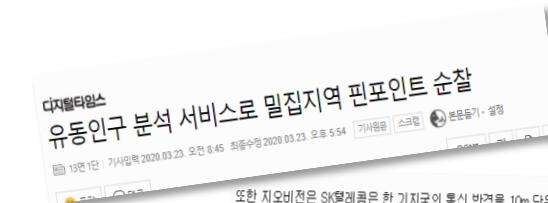
[의정부시청] 코로나19 사회적 거리두기 안내
▲ 현재 (의정부 제일시장)의 인구 '혼잡'
코로나 예방을 위해서 기급적 이동을 자제해주시기
바랍니다.



[경기도청] 코로나19의 감염 확산 방지를 위해
주민 여러분들의 자발적인 참여와 실천이
중요합니다.
▲ 경기도가 인구밀집 지도를 제공합니다.
<http://bit.ly/covid19>

[실시간 유동인구 파악 기술 도입 사례]

사례1) '지오비전' 기반 유동인구 데이터로
경북 경찰청 핀포인트 순찰 지원



또한 지오비전은 SK텔레콤은 한 기지국의 통신 번호를 10m 단위로 갈고 꾸개 관리하는
'피셀' 기술을 적용해 기존 '셀 ID 측위 방식' 대비 5배 이상 정밀한 위치 정보를 제공할 수
있다. 경북경찰청은 지오비전 데이터를 활용해 현재 인구 밀집도가 높은 지역을 확인, 실
시간 '핀 포인트' 순찰을 시행해 코로나19 확산을 막기하고 주민의 체감안전도 향상에 나
설 계획이다.

사례2) 경기 성남시 실시간 유동인구 분석 기술 도입
인구 밀집도가 높은 지역을 중심으로
코로나19 방역 대책 마련 및 소독 경로 측정

출처 및 참고자료

구분	출처 기관	URL
변수 데이터	경기도청	https://www.gg.go.kr
	경기도 감염병관리지원단	http://www.gidcc.or.kr
참고 기사	동아사이언스 (2차 공격률 정보 제공)	https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=103&oid=584&aid=0000009104
	경향신문 (성남시 사례)	https://m.khan.co.kr/view.html?art_id=202005111508021#c2b
	디지털 타임스 (경북경찰청 사례)	https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=105&oid=029&aid=0002589934
	연합뉴스 (사회적 거리두기 사진)	https://www.yna.co.kr/view/AKR20200628043400530

활용 프로그램



THANK YOU
코로나 종식을 응원합니다!