

2023

화성특례시 도약을 위한 인구 시계열 예측 연구

| 결과보고서 |



내 삶을 바꾸는
희망 화성

| 집필 |

화성시청 스마트도시과

기획조정실장	정승호
스마트도시과장	김명숙
빅데이터팀장	신환철
주 무 관	이동재

2023. 6.



목 차

I 서 론

○ 시작말	6
-------------	---

II 인구예측 분석

1. 데이터 전처리	8
2. 단위근 검정	16
3. 시계열 분석모델 성능비교	21

III 인구예측 결과

○ 화성시 100만 인구도달 시기	26
--------------------------	----

IV 참고문헌	29
---------------	----

요 약

연구모델

□ 예측변수 선정

이번 프로젝트에서는 주거환경과 국내경제상황이 화성시 거주인구에 영향을 미칠 것으로 가정하고 연구를 진행하였다. 화성시 인구에 영향을 미칠 요인으로 ▲토지가격 ▲아파트 매매 · 전세 · 월세 지수 ▲전입 · 전출 ▲출생 · 사망 ▲전산업생산지수 ▲CD금리(은행이 발행하는 양도성예금증서의 금리) ▲CP금리(기업이 자금조달을 목적으로 발행하는 무담보 기업어음) ▲국고채 금리 ▲회사채 금리 ▲기준금리 ▲실업률 ▲소비자물가지수 등 106개 종류의 설명변수 데이터를 수집하였다.

□ 분석모델 성능비교

이번 프로젝트의 분석모델은 머신러닝 모형활용 예측보고서(장나원 외, 2022)에서 활용한 모델인 다음과 같은 총 16개의 시계열 모형과 머신러닝 방법들을 고려하였다.

- ① RandomWalk ② AR ③ LASSO ④ Adaptive LASSO ⑤ Elastic Net ⑥ Adaptive Elastic Net
- ⑦ Ridge regression ⑧ CSR(complete subset regression) ⑨ Target Factor ⑩ Random Forest
- ⑪ RF / OLS ⑫ AdaLasso / RF ⑬ XGBoost ⑭ NN(Neural Network) ⑮ LSTM
- ⑯ Boruta_selected_RF 총 16개 모델의 예측력을 비교분석하였다.

□ 분석모델 선택

모든 개월에서 AR모델이 좋은 예측성능을 보였으며, Random Forest와 AdaLasso/RF 모델도 각각 7개월 예측과 10개월 예측을 제외한 기간에서 우수한 예측성능을 보였다. 이에 따라, 이번 프로젝트에서는 AR모델을 활용하여 인구추이를 예측하였다.

요 약

예측결과



예측결과, 화성시 인구 100만 도달시기는 2023년 9월과 10월 사이가 될 것으로 예상된다. 시계열예측은 지금까지의 데이터로 미래 데이터를 예측하기, 지금까지 없었던 변수가 생길 경우 시계열흐름은 요동칠 수 있다. 예를 들어, 타 지자체 거주 화성시 공무원들에게 화성시로의 전입 시 인센티브를 주는 정책을 펼치거나, 타 지자체보다 급격히 아동양육 지원금액을 대폭 향상시키거나, 외국인 근로자 채용 지원금 확대 정책시행 등 지금까지 없었던 변수가 발생하게 되면, 이에따라 시계열 예측결과는 다소 빗겨갈 수 있음을 인지해야한다.

〈표〉 화성시 인구추이 예측치

순번	연월	예측 인구성장률	총인구수(명)
1	2023년 4월	-	978,653
2	2023년 5월	0.0038899	982,460
3	2023년 6월	0.0041965	986,583
4	2023년 7월	0.0048495	991,367
5	2023년 8월	0.0044766	995,805
6	2023년 9월	0.0033211	999,112
7	2023년 10월	0.0028119	1,001,922
8	2023년 11월	0.0027496	1,004,677
9	2023년 12월	0.0023012	1,006,989
10	2024년 1월	0.0025762	1,009,583
11	2024년 2월	0.0026093	1,012,217
12	2024년 3월	0.0022981	1,014,543
13	2024년 4월	0.0025841	1,017,165

I. 서론

□ 인구, 모든 정책의 시작이자 끝

한국은행 이창용 총재는 올해 5월 25일에 열린 “통화정책방향 관련 총재 기자간담회”에서 대한민국 올해 경제성장률을 1.4%로 하향조정하면서, 이같은 저성장 국면을 극복하기 위해서는 단순히 통화정책과 재정정책으로 해결하려는 것이 아닌, 우리 사회구조의 전반적인 개혁이 필요하다고 말했다. 또 대한민국은 이미 장기저성장 구조에 와있다면서, 그 이유로 저출산(저출생)과 고령화를 꼬집었다.

경제유튜버 슈카월드(서울대 경제학과)는 이와 관련하여 일본과 한국을 비교하며, 한은 총재가 얘기했던 장기저성장을 설명하였다. 슈카월드는 “일본의 경우 소위 ‘잃어버린 30년’이라 표현되는 장기저성장의 대표국가이다. 30년 간 일본의 경제성장률은 평균 약 1%대를 기록하였고, 월급이 그대로인 나라이다. 장기저성장의 이유로 제시되는게 플라자합의, 미일 반도체협정 등 경제적 충격이 강하게 왔을 때 그것을 극복하지 못했다는 것이다. 강한 경제적 충격을 극복하지 못한 이유로 많은 것이 제시되지만, 가장 심각한 것은 일본의 생산가능인구가 1995년에 고점을 찍고, 그 이후로 계속 하락추세이기 때문이다.” 라고 했다.

이를 한국상황에 비교하며 “한국도 잠재성장률 추가 하락 및 장기저성장의 위험에 와있다. 2020년 대 이후로 생산가능 인구가 감소추세로 접어들었기 때문에, 경제충격에서의 회복력이 이전보다 상당히 약화되어있다. 2023년 1분기 합계출산율은 0.81명으로 1분기 기준 역대 최저치를 기록했으며, 전년 동기 대비 6%(4,116명) 감소한 수치이다.” 라고 하며 한국이 장기저성장 국면에서 회복하지 못할 가능성에 대해 염려하였다. 이같은 경제전문가들의 염려처럼, 한국이 경제적으로 회복탄력성을 갖추기 위해서는 그 무엇보다도 생산가능인구를 확보하는 것이 가장 중요하다.

□ 연구목적 및 방법

이번 프로젝트에서는 이렇듯 사회기반에서 가장 중요한 분야인 **인구**에 대한 연구를 진행하였다. 그 중에서도 화성시의 인구추이를 시계열분석으로 예측해보고자하였다. 곧 100만을 앞둔 화성시의 인구추이에측 시계열모델을 구축하여, 향후 화성시의 인구추이를 지속적으로 예측해 관련 정책들을 사전에 마련할 수 있는 근거를 제시하고자 한다.

통계청에서는 코호트요인법을 활용하여 ‘장래인구추계’를 발표하고 있다. 그러나, 「한국의 시도별 장래인구 예측」 보고서¹⁾의 저자는 “전통적 인구예측방법인 코호트요인법은 코호트별 기간 변동률의 가정에 의해 인구예측의 정확성을 더 기할 수 있지만, 코호트별 세분됨으로써 변동요인에 의한 추정에 어려움이 증가한다. 따라서 코호트요인법보다는 최근의 시계열 변화를 더 중시한, 인구 및 경제의 시계열 모형을 이용한 확률적 인구예측방법이 간단하고도 효율적일 것으로 여겨진다.” 라고 했다. 이에 근거하여, 다양한 시계열 모형과 머신러닝 모델을 활용하여 화성시의 인구추이를 예측해보고자 한다.

이 결과보고서에서 수정되어야하는 부분이 있거나, 다른 모델로 더 좋은 예측성능을 도출해내었다면, 화성시청 스마트도시과 빅데이터팀 이동재 주무관(aisw@korea.kr)에게 연락주길 희망한다.

1) 김형기, 문경중, “한국의 시도별 장래인구 예측”, 국토계획 제46권 제6호(2011): 80.

II. 인구예측 분석

1. 데이터 수집 및 전처리

□ 예측변수 선정

이번 프로젝트에서는 주거환경과 국내경제상황이 화성시 거주인구에 영향을 미칠 것으로 판단하였다. 화성시 인구에 영향을 미칠 요인으로 토지가격, 아파트 매매·전세·월세 지수, 전입·전출, 출생·사망, 전산업생산지수, CD금리(은행이 발행하는 양도성예금증서의 금리), CP금리(기업이 자금조달을 목적으로 발행하는 무담보 기업어음), 국고채 금리, 회사채 금리, 기준금리, 실업률, 소비자물가지수 등 106개 종류의 설명변수 데이터를 수집하였다.

□ 데이터 출처

순번	종류	출처
1	내국인데이터	차세대 주민등록 통합행정시스템
2	외국인데이터	법무부 외국인정보공동이용시스템
3	토지가격·아파트 가격지수	한국부동산원
4	전입·전출	KOSIS
5	출생자수, 사망자수	행정안전부 주민등록 인구통계
6	택지개발사업, 공공주택지구 조성사업	화성시 신도지재생과
7	도시개발사업, 복합도시개발사업	화성시 지역개발과
8	전산업생산지수, 시장금리, 기준금리, 실업률, 소비자물가지수	한국은행 통계정보시스템

□ 결측치 확인

106개의 변수중 결측치가 있는 변수는 총 3개의 종류의 변수로 확인하였다. 2015년 6월부터 조사를 시작한 월세가격지수, 월세가격변동률가 각각 17개, 18개의 결측치가 존재하였으며, 전산업생산지수(IAIP, 'PI' 로 통칭)의 경우 23년 4월 1개의 결측치가 존재하였다.

```
Python 코드
import pandas as pd
import os

# 모든 행, 열 표시
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

# 파일읽기
df = pd.read_excel("hspop_v2.xlsx")

# 결측치 확인하기
df.isna().sum()
```


순번	변수명	결측치
1	총인구_남	0
2	총인구_여	0
3	내국인	0
4	내국인_남	0
5	내국인_여	0
6	외국인	0
7	외국인_남	0
8	외국인_여	0
9	등록외국인	0
10	등록외국인_남	0
11	등록외국인_여	0
12	외국국적동포	0
13	외국국적동포_남	0
14	외국국적동포_여	0
15	지가지수	0
16	지가변동률_전국	0
17	지가변동률_수도권	0
18	지가변동률_지방	0
19	지가변동률_대도시	0
20	지가변동률_시지역	0
21	지가변동률_군지역	0
22	지가변동률_화성시	0
23	매매가격지수_아파트_전국	0
24	매매가격지수_아파트_수도권	0
25	매매가격지수_아파트_지방권	0
26	매매가격지수_아파트_6대광역시	0
27	매매가격지수_아파트_5대광역시	0
28	매매가격지수_아파트_9개도	0
29	매매가격지수_아파트_8개도	0
30	매매가격지수_아파트_서울	0
31	매매가격지수_아파트_화성시	0
32	매매가격변동률_아파트_전국	0
33	매매가격변동률_아파트_수도권	0
34	매매가격변동률_아파트_지방권	0
35	매매가격변동률_아파트_6대광역시	0
36	매매가격변동률_아파트_5대광역시	0
37	매매가격변동률_아파트_9개도	0
38	매매가격변동률_아파트_8개도	0
39	매매가격변동률_아파트_서울	0
40	매매가격변동률_아파트_화성시	0
41	전세가격지수_아파트_전국	0
42	전세가격지수_아파트_수도권	0
43	전세가격지수_아파트_지방권	0
44	전세가격지수_아파트_6대광역시	0
45	전세가격지수_아파트_5대광역시	0
46	전세가격지수_아파트_9개도	0
47	전세가격지수_아파트_8개도	0
48	전세가격지수_아파트_서울	0
49	전세가격지수_아파트_화성시	0
50	전세가격변동률_아파트_전국	0
51	전세가격변동률_아파트_수도권	0
52	전세가격변동률_아파트_지방권	0

53	전세가격변동률_아파트_6대광역시	0
54	전세가격변동률_아파트_5대광역시	0
55	전세가격변동률_아파트_9개도	0
56	전세가격변동률_아파트_8개도	0
57	전세가격변동률_아파트_서울	0
58	전세가격변동률_아파트_화성시	0
59	월세가격지수_아파트_전국	17
60	월세가격지수_아파트_수도권	17
61	월세가격지수_아파트_지방권	17
62	월세가격지수_아파트_6대광역시	17
63	월세가격지수_아파트_5대광역시	17
64	월세가격지수_아파트_9개도	17
65	월세가격지수_아파트_8개도	17
66	월세가격지수_아파트_서울	17
67	월세가격지수_아파트_화성시	17
68	월세가격변동률_아파트_전국	18
69	월세가격변동률_아파트_수도권	18
70	월세가격변동률_아파트_지방권	18
71	월세가격변동률_아파트_6대광역시	18
72	월세가격변동률_아파트_5대광역시	18
73	월세가격변동률_아파트_9개도	18
74	월세가격변동률_아파트_8개도	18
75	월세가격변동률_아파트_서울	18
76	월세가격변동률_아파트_화성시	18
77	총전입	0
78	총전출	0
79	순이동	0
80	출생자수	0
81	출생자수_남자	0
82	출생자수_여자	0
83	사망자수	0
84	사망자수_남자	0
85	사망자수_여자	0
86	PI_P(전산업생산지수-광공업)	1
87	PI_C(전산업생산지수-건설업)	1
88	PI_SI(전산업생산지수-서비스업)	1
89	PI_PA(전산업생산지수-공공행정)	1
90	CD(CD금리91일)	0
91	CP(CP금리91일)	0
92	YT_1(국고채1년)	0
93	YT_3(국고채3년)	0
94	YT_5(국고채5년)	0
95	YT_10(국고채10년)	0
96	CB_3_AA(회사채 3년, AA-)	0
97	CB_3_BBB(회사채 3년, BBB-)	0
98	BaseRate(한국은행 기준금리)	0
99	실업률	0
100	CPI_F(소비자물가지수-식료품 및 비주류음료)	0
101	CPI_C(소비자물가지수-의류 및 신발)	0
102	CPI_H(소비자물가지수-가정용품 및 가사서비스)	0
103	CPI_HE(소비자물가지수-보건)	0
104	CPI_T(소비자물가지수-교통)	0
105	CPI_E(소비자물가지수-교육)	0
106	CPI_M(소비자물가지수-기타상품 및 서비스)	0

□ 시계열 결측치 처리방법(5가지)

1) Forward Filling

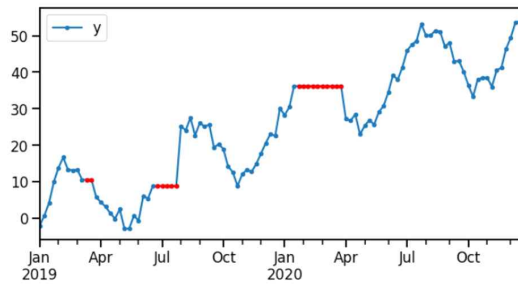
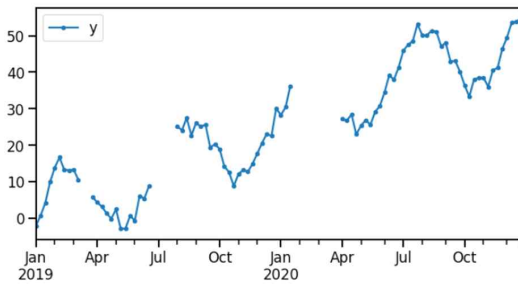
결측값 발생 이전의 가장 마지막 값으로 이후의 결측값을 대체하는 방법

Python 코드
df_forward = df.fillna(method="ffill")

Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	nan
2020-01-05	nan
2020-01-06	nan
2020-01-07	58
2020-01-08	5



Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	23
2020-01-05	23
2020-01-06	23
2020-01-07	58
2020-01-08	5



2) Backward Filling

결측값 발생 이후의 가장 첫 번째 값으로 결측값을 대체하는 방법

Python 코드
df_backward = df.fillna(method="bfill")

Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	nan
2020-01-05	nan
2020-01-06	nan
2020-01-07	58
2020-01-08	5



Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	58
2020-01-05	58
2020-01-06	58
2020-01-07	58
2020-01-08	5

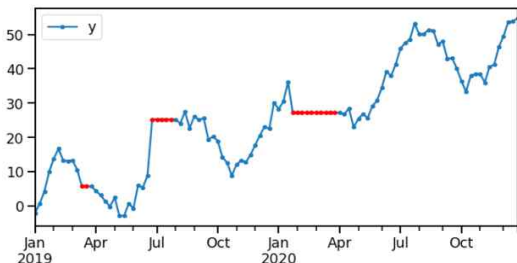
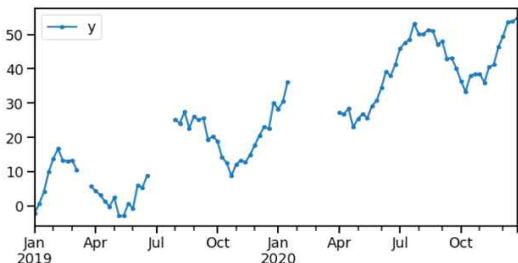


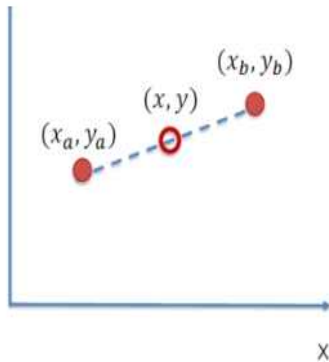
그림 : Kaggle Time Series Study

3) Linear Interpolation

결측값 발생 이전과 이후의 데이터의 선형결합으로 결측값을 대체하는 방법

```
Python 코드
# 시간간격이 일정한 경우
df_interpolate = df.interpolate(method='linear')

# 시간간격이 불규칙한 경우
df_interpolate = df.interpolate(method='time')
```



$$y = y_a + (y_b - y_a) \frac{x - x_a}{x_b - x_a} \text{ at the point } (x, y)$$

Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	nan
2020-01-05	nan
2020-01-06	nan
2020-01-07	58
2020-01-08	5



Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	31.75
2020-01-05	40.50
2020-01-06	49.25
2020-01-07	58
2020-01-08	5

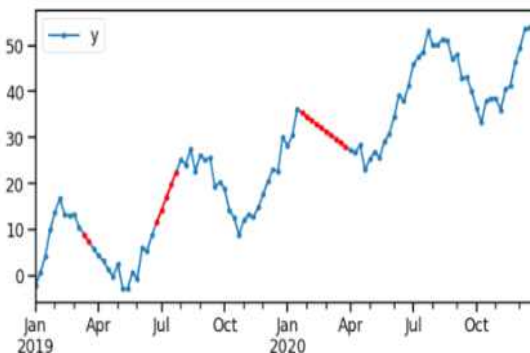
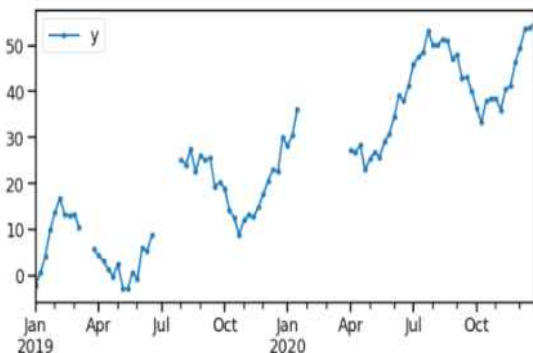


그림 : Kaggle Time Series Study

시계열 데이터가 위 그림처럼 상향 또는 하향의 추세가 있는 데이터일 경우, Forward와 Backward Filling 방법보다 Linear Interpolation 방법이 더 적합하다.

4) Spline Interpolation

다항식을 사용하여 결측치 보완을 부드럽게 연결하는 방법

Python 코드

```
df_interpolate = df.interpolate(method='spline', order='다항식의 차수')
```

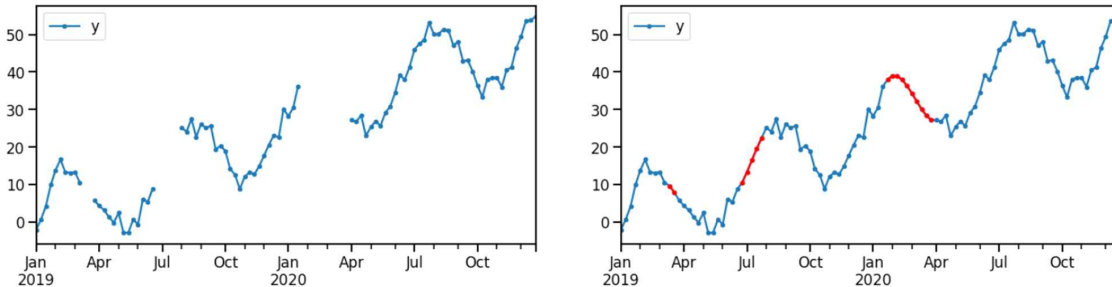


그림 : Kaggle Time Series Study

결측치의 개수가 많아도 앞서 설명한 방법에 비해 시계열데이터의 왜곡이 적다는 장점이 있으나, 계산비용이 더 크다는 단점이 있다.

5) STL(Season-Trend decomposition using LOESS) and Interpolation

Python 코드

라이브러리 불러오기

```
from statsmodels.datasets import co2
```

```
import matplotlib.pyplot as plt
```

```
from pandas.plotting import register_matplotlib_converters
```

```
register_matplotlib_converters()
```

```
from statsmodels.tsa.seasonal import STL
```

```
# (1) linear interpolate방법으로 결측치 대체(ffill, bfill 모두 가능) 후 STL 모델학습
res = STL(df.interpolate(method='linear')).fit()
```

```
# (2) 계절성 요소 추출
```

```
seasonal_component = res.seasonal
```

```
# (3) 계절성 제거
```

```
df_deseasonalised = df['y']-seasonal_component
```

```
# (4) 계절성 제거된 데이터에서 선형보간(linear interpolation)
```

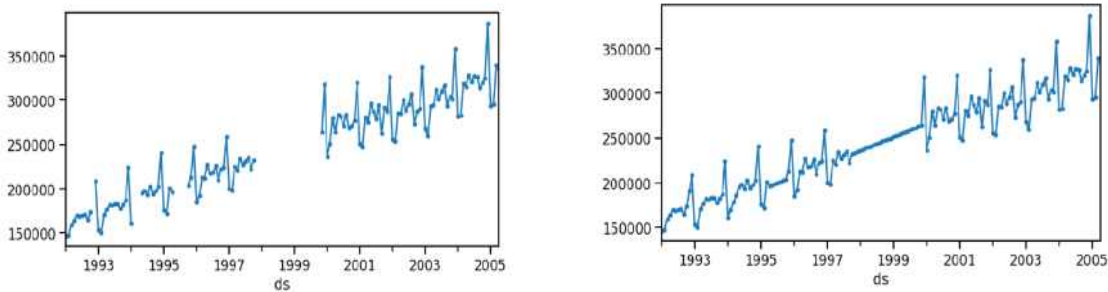
```
df_deseasonalised_imputed = df_deseasonalised.interpolate(method='linear')
```

```
# (5) 계절성 요소를 다시 더한 후 보간 종료
```

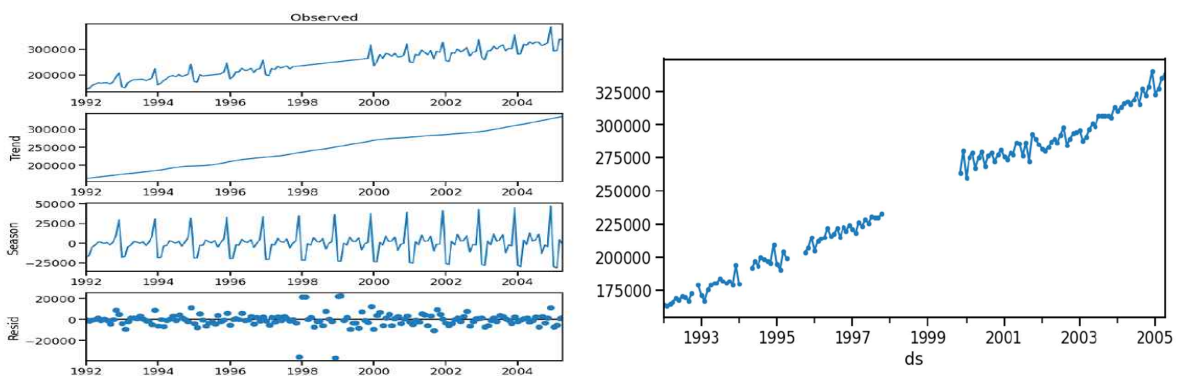
```
df_imputed = df_deseasonalised_imputed + seasonal_component
```

5) STL(Season-Trend decomposition using LOESS) and Interpolation

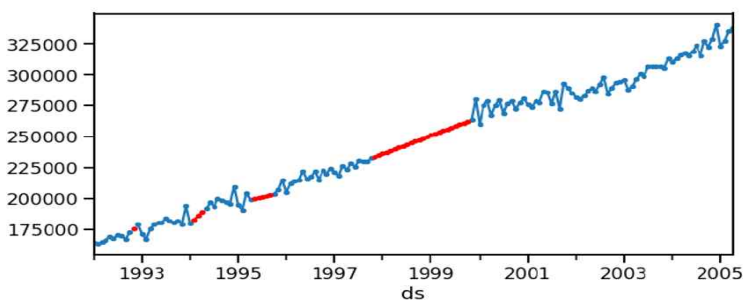
① 앞서 설명한 방법(ffill, bfill, linear interpolation 등)으로 결측값을 대체한다



② 추세(trend), 계절성(seasonality), 잡음(noise)으로 분해한 뒤, 계절성 부분을 제거한다.



③ 계절적 요소가 제거된 데이터에 대해서 여러 결측값 대체 방법을 사용하여 결측값을 대체한다.



④ 계절적인 요소를 다시 추가한다

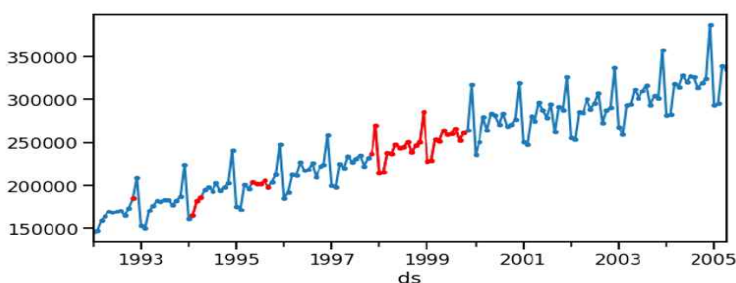


그림 : Kaggle Time Series Study

□ 결측치 처리방법 선택기준

- 연속적인 결측값 간격이 작을 경우 : Forward filling, Linear filling
- 연속적인 결측값 간격이 크고 추세(trend), 계절성(seasonality)가 없는 경우 : Forward filling, Linear filling, Mean
- 연속적인 결측값 간격이 크고 추세(trend)는 강하나, 계절성(seasonality)이 없는 경우 : Linear interpolation 혹은 Spline interpolation
- 연속적인 결측값 간격이 크고 계절성(seasonality)이 있는 경우 : STL and interpolation

□ 결측치 처리방법 선택

- 연속적인 결측값의 간격이 작으므로, 앞부분의 데이터가 없는 월세가격데이터는 Backward Filling 방법으로, 뒷부분의 데이터가 없는 전산업생산지수는 Forward Filling 방법으로 결측치를 처리하였다.

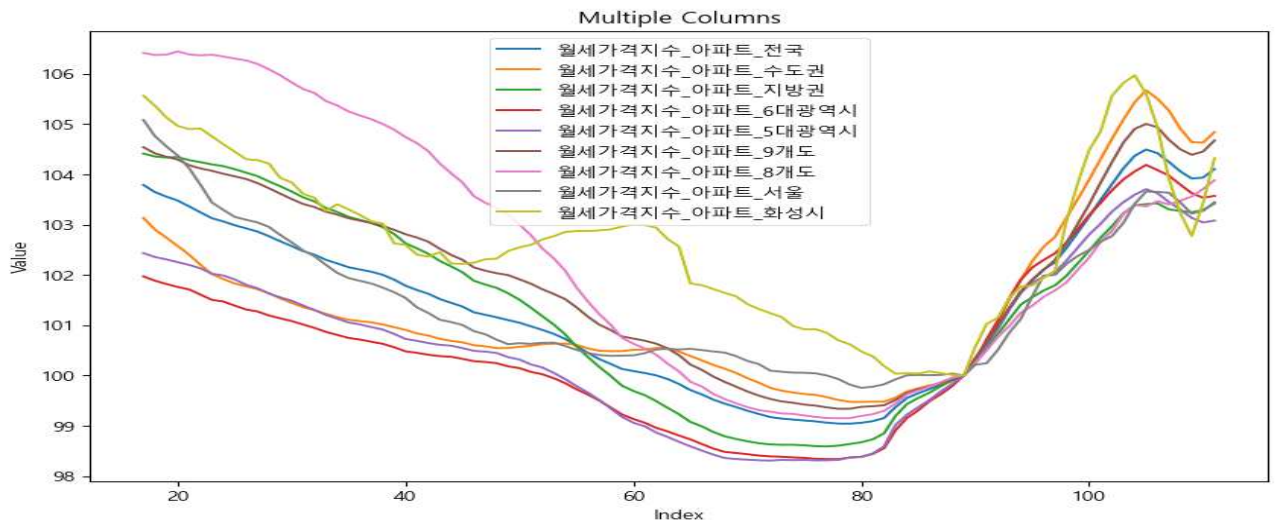


그림 : 아파트 월세가격지수 추이

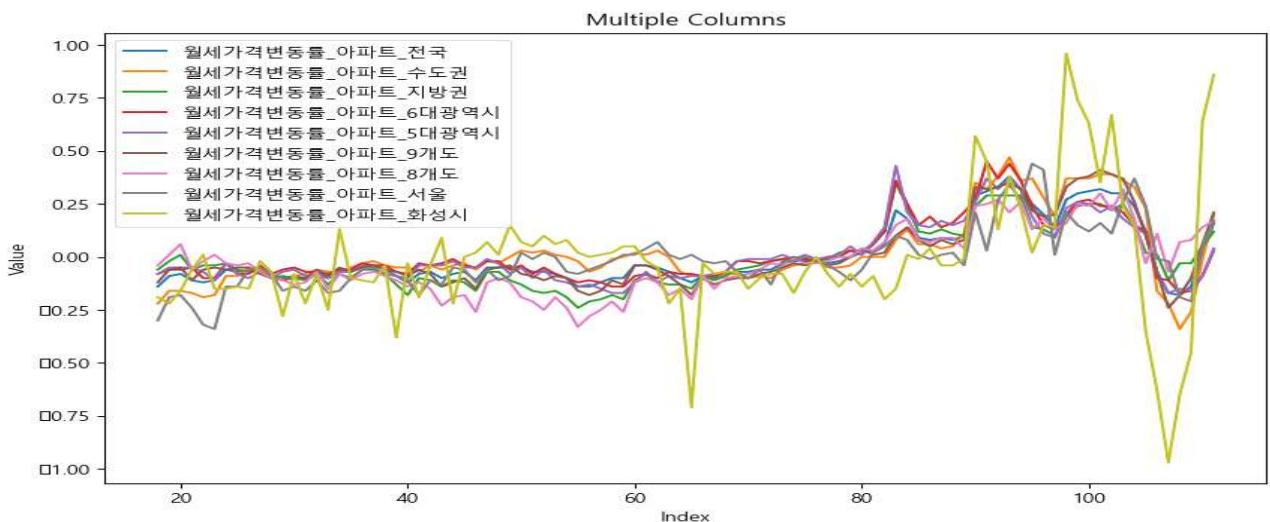


그림 : 아파트 월세가격변동률 추이

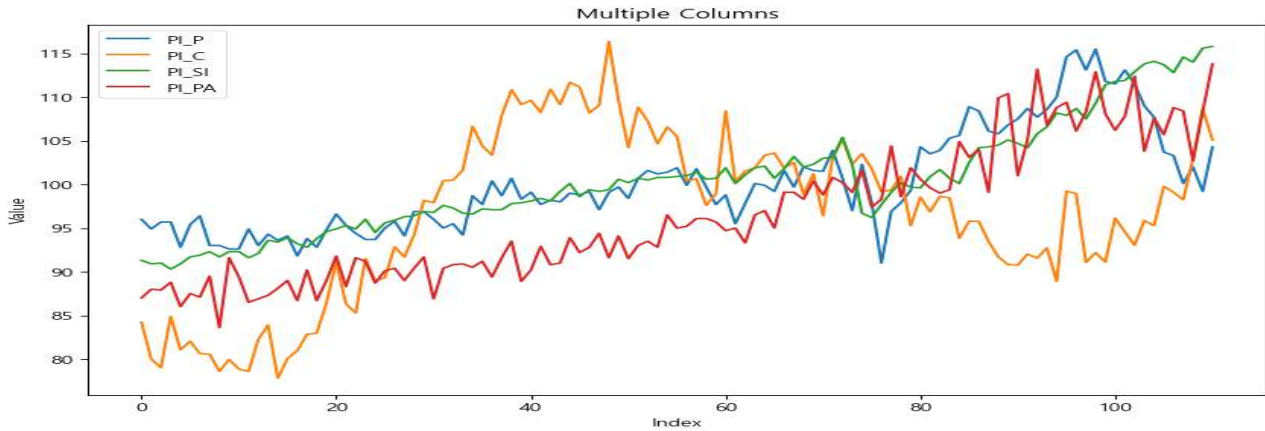


그림 : 전산업생산지수 분야별 추이

2. 단위근 검정(UnitRoot Test)

○ 정상성

단위근에 대해 설명하기 위해서는, 먼저 정상성의 개념을 알아야한다. ‘정상성(Stationarity)’이란 시간에 상관없이 일정한 통계적 성질(평균, 분산, 공분산)을 가지고 있는 것을 말한다. 정상성을 가진 시계열 데이터는 시간에 따라 추세가 없으며, 공분산이 특정 시점에 의존하지 않고, 시차에 의존한다.

○ 단위근

시계열데이터가 ‘정상성’을 띠면, ‘단위근이 존재하지 않는다’라고 하며, 비정상성을 띠면, ‘단위근이 존재한다’라고 한다. 시계열 분석을 위해서는 단위근을 제거하여 시계열 데이터가 일정한 통계적 성질을 가지도록 해야한다.

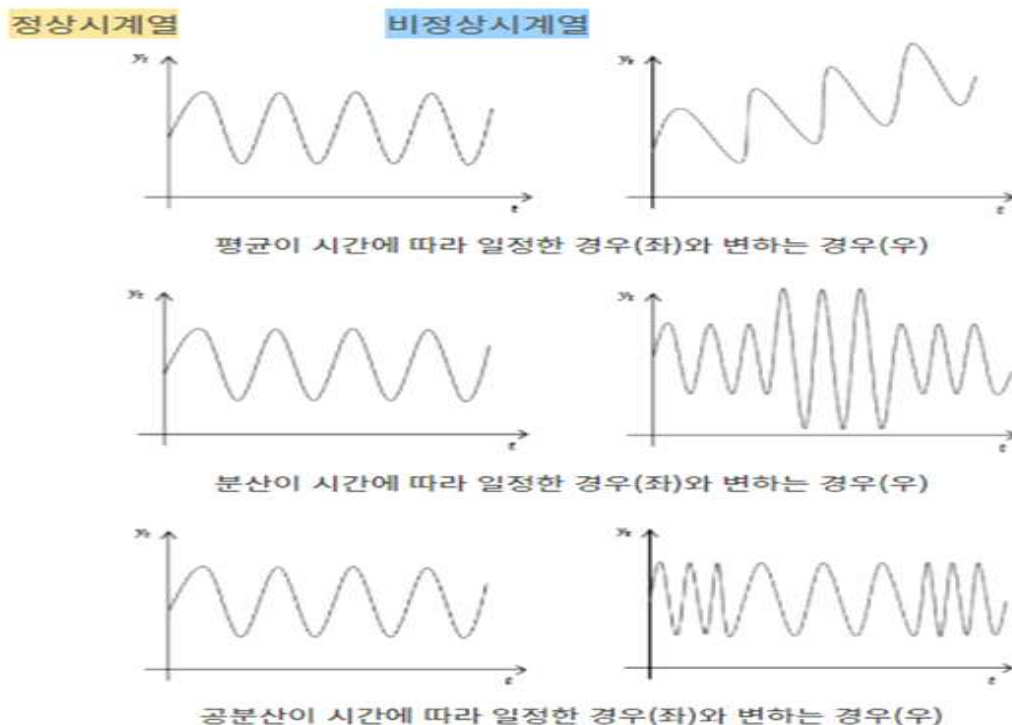


그림 : Kaggle Time Series Study

□ 단위근 검정방법

단위근 검정방법에는 크게 ADF, PP, KPSS Test 총 3가지의 방법이 있다. ADF와 PP는 귀무가설이 단위근이 존재한다(비정상시계열이다)이고, KPSS의 귀무가설은 단위근이 존재하지 않는다(정상성 시계열이다)이다. 통계량(Statistics)의 절대값이 기각역의 절대값보다 작은 경우 귀무가설을 채택하고, 반대의 경우 귀무가설을 기각한다. 예를 들어, ADF Test 통계량의 절대값이 기각역의 절대값보다 작은 경우, 귀무가설을 채택하게 되고, 이 경우 시계열 데이터는 단위근이 존재한다는 것이다. 즉, 비정상시계열 데이터이므로, 단위근 제거가 필요하다는 의미이다.

1. ADF(Augmented Dicky-Fuller) Test, PP(Phillips-Perron) Test

- ▲ 귀무가설 = 단위근이 존재한다(비정상성 시계열)
- ▲ 대립가설 = 단위근이 존재하지 않는다(정상성 시계열)

2. KPSS(Kwiatkowski-Phillips-Schmidt-Shin) Test

- ▲ 귀무가설 = 단위근이 존재하지 않는다(정상성 시계열)
- ▲ 대립가설 = 단위근이 존재한다(비정상성 시계열)

□ 단위근 제거

단위근 검정결과 비정상성으로 판단된 시계열데이터는 정상성 데이터로 변환하는 과정이다. 아래 표의 Tcode는 단위근을 제거하는 방법의 종류를 나타낸다. 예를 들어, Tcode 5 방식으로 단위근을 제거한다는 것은 원본 데이터에 로그를 취하고 차분(diff)를 진행하여 정상성데이터로 만드는 과정을 말한다.

Tcode	변환방법
1	변환 없음
2	diff(x)
3	diff(diff(x))
4	log(x)
5	diff(log(x))
6	diff(diff(log(x)))

<표> 단위근 검정 및 Tcode

순번	변수명	ADF Test 통계량 (tau2)	ADF 기각역 (0.05)	귀무가설	KPSS Test 통계량	KPSS 기각역 (0.05)	귀무가설	Tcode
1	총인구	-1.00	-3.43	채택	0.17	0.14	기각	5
2	총인구_남							5
3	총인구_여							5
4	내국인							5
5	내국인_남							5
6	내국인_여							5
7	외국인							5
8	외국인_남							5
9	외국인_여							5
10	등록외국인							5
11	등록외국인_남							5
12	등록외국인_여							5
13	외국국적동포							5
14	외국국적동포_남							5
15	외국국적동포_여							5
16	지가지수	-0.14	-2.88	채택	0.95	0.46	기각	3
17	지가변동률_전국	-1.86	-2.88	채택	0.20	0.14	기각	2
18	지가변동률_수도권	-2.04	-2.88	채택	0.32	0.46	채택	2
19	지가변동률_지방	-1.56	-2.88	채택	0.21	0.46	채택	2
20	지가변동률_대도시	-2.13	-2.88	채택	0.23	0.46	채택	2
21	지가변동률_시지역	-1.87	-2.88	채택	0.27	0.46	채택	2
22	지가변동률_군지역	-1.58	-2.88	채택	0.20	0.46	채택	2
23	지가변동률_화성시	-3.22	-2.88	기각	0.33	0.46	채택	1
24	매매가격지수_아파트_전국							3
25	매매가격지수_아파트_수도권							3
26	매매가격지수_아파트_지방권							3
27	매매가격지수_아파트_6대광역시							3
28	매매가격지수_아파트_5대광역시							3
29	매매가격지수_아파트_9개도							3
30	매매가격지수_아파트_8개도							3
31	매매가격지수_아파트_서울							3
32	매매가격지수_아파트_화성시							3
33	매매가격변동률_아파트_전국	-2.50	-2.88	채택	0.10	0.46	채택	2
34	매매가격변동률_아파트_수도권	-2.70	-2.88	채택	0.11	0.46	채택	2
35	매매가격변동률_아파트_지방권	-2.33	-2.88	채택	0.09	0.46	채택	2
36	매매가격변동률_아파트_6대광역시	-2.20	-2.88	채택	0.11	0.46	채택	2
37	매매가격변동률_아파트_5대광역시	-2.18	-2.88	채택	0.13	0.46	채택	2
38	매매가격변동률_아파트_9개도	-2.38	-2.88	채택	0.08	0.46	채택	2
39	매매가격변동률_아파트_8개도	-2.28	-2.88	채택	0.09	0.46	채택	2
40	매매가격변동률_아파트_서울	-3.78	-2.88	기각	0.29	0.46	채택	1
41	매매가격변동률_아파트_화성시	-3.22	-2.88	기각	0.11	0.46	채택	1
42	전세가격지수_아파트_전국							2
43	전세가격지수_아파트_수도권							2
44	전세가격지수_아파트_지방권							2

45	전세가격지수_아파트_6대광역시							2
46	전세가격지수_아파트_5대광역시							2
47	전세가격지수_아파트_9개도							2
48	전세가격지수_아파트_8개도							2
49	전세가격지수_아파트_서울							2
50	전세가격지수_아파트_화성시							2
51	전세가격변동률_아파트_전국	-3.66	-2.88	기각	0.22	0.46	채택	1
52	전세가격변동률_아파트_수도권	-4.23	-2.88	기각	0.29	0.46	채택	1
53	전세가격변동률_아파트_지방권	-2.65	-2.88	채택	0.13	0.46	채택	2
54	전세가격변동률_아파트_6대광역시	-2.70	-2.88	채택	0.20	0.46	채택	2
55	전세가격변동률_아파트_5대광역시	-2.51	-2.88	채택	0.19	0.46	채택	2
56	전세가격변동률_아파트_9개도	-3.56	-2.88	기각	0.20	0.46	채택	1
57	전세가격변동률_아파트_8개도	-2.40	-2.88	채택	0.10	0.46	채택	2
58	전세가격변동률_아파트_서울	-5.07	-2.88	기각	0.38	0.46	채택	1
59	전세가격변동률_아파트_화성시	-3.97	-2.88	기각	0.25	0.46	채택	1
60	월세가격지수_아파트_전국	-0.98	-3.43	채택	0.20	0.14	기각	3
61	월세가격지수_아파트_수도권							3
62	월세가격지수_아파트_지방권							3
63	월세가격지수_아파트_6대광역시							3
64	월세가격지수_아파트_5대광역시							3
65	월세가격지수_아파트_9개도							3
66	월세가격지수_아파트_8개도							3
67	월세가격지수_아파트_서울							3
68	월세가격지수_아파트_화성시							3
69	월세가격변동률_아파트_전국	-2.53	-2.88	채택	0.62	0.46	기각	2
70	월세가격변동률_아파트_수도권	-2.59	-2.88	채택	0.63	0.46	기각	2
71	월세가격변동률_아파트_지방권	-1.98	-2.88	채택	0.49	0.46	기각	2
72	월세가격변동률_아파트_6대광역시	-2.21	-2.88	채택	0.48	0.46	기각	2
73	월세가격변동률_아파트_5대광역시	-2.35	-2.88	채택	0.49	0.46	기각	2
74	월세가격변동률_아파트_9개도	-2.51	-2.88	채택	0.56	0.46	기각	2
75	월세가격변동률_아파트_8개도	-1.34	-2.88	채택	0.44	0.46	채택	2
76	월세가격변동률_아파트_서울	-2.39	-2.88	채택	0.81	0.46	기각	2
77	월세가격변동률_아파트_화성시	-3.50	-2.88	기각	0.44	0.46	채택	1
78	총전입							5
79	총전출							5
80	순이동							5
81	출생자수							2
82	출생자수_남자							2
83	출생자수_여자							2
84	사망자수							2
85	사망자수_남자							2
86	사망자수_여자							2
87	PI_P(전산업생산지수-광공업)							5
88	PI_C(전산업생산지수-건설업)							5
89	PI_SI(전산업생산지수-서비스업)							5
90	PI_PA(전산업생산지수-공공행정)							5
91	CD(CD금리91일)							2
92	CP(CP금리91일)							2

93	YT_1(국고채1년)							2
94	YT_3(국고채3년)							2
95	YT_5(국고채5년)							2
96	YT_10(국고채10년)							2
97	CB_3_AA(회사채 3년, AA-)							2
98	CB_3_BBB(회사채 3년, BBB-)							2
99	BaseRate(한국은행 기준금리)							2
100	실업률							2
101	CPI_F(소비자물가지수-식료품 및 비주류음료)							5
102	CPI_C(소비자물가지수-의류 및 신발)							5
103	CPI_H(소비자물가지수-가정용품 및 가사서비스)							5
104	CPI_HE(소비자물가지수-보건)							5
105	CPI_T(소비자물가지수-교통)							5
106	CPI_E(소비자물가지수-교육)							5
107	CPI_M(소비자물가지수-기타상품 및 서비스)							5

3. 시계열 예측모델 성능비교

□ 설명변수

이번 프로젝트에서는 「머신러닝 방법을 활용한 한국의 수출입 증가율 예측 및 분석」(장나원, 한희준, 2022) 논문과 동일한 개수의 설명변수로 시계열 예측을 진행하였다. 이 논문에서는 “52개의 변수와 52개 변수에서 추출한 4개의 주성분요소(principal component factor)들의 1차에서 4차까지의 시차변수들과 종속변수인 총인구수의 1차에서 4차까지의 시차변수 등을 개별 설명변수로 취급하여, 각 시계열 모형 및 머신러닝 모형에 사용되는 설명변수는 총 228개($(52+4) \times 4 + 4$)로 설정하였다.”²⁾ 라고 설명변수의 개수를 설명하였다.

○ 이번 프로젝트에서 사용된 설명변수

R 코드		
Y = Y[,c(1:15 ,23, 32,41,50,59,68,77:107)]		
연번	기존번호	변수명
1	1	총인구
2	2	총인구_남
3	3	총인구_여
4	4	내국인
5	5	내국인_남
6	6	내국인_여
7	7	외국인
8	8	외국인_남
9	9	외국인_여
10	10	등록외국인
11	11	등록외국인_남
12	12	등록외국인_여
13	13	외국국적동포
14	14	외국국적동포_남
15	15	외국국적동포_여
16	23	지가변동률_화성시
17	32	매매가격지수_아파트_화성시
18	41	매매가격변동률_아파트_화성시
19	50	전세가격지수_아파트_화성시
20	59	전세가격변동률_아파트_화성시
21	68	월세가격지수_아파트_화성시
22	77	월세가격변동률_아파트_화성시
23	78	총전입
24	79	총전출
25	80	순이동
26	81	출생자수
27	82	출생자수_남자
28	83	출생자수_여자
29	84	사망자수

2) 장나원, 한희준, “머신러닝 방법을 활용한 한국의 수출입 증가율 예측 및 분석”, 국제경제연구 제28권 제4호, 2022, 61쪽.

30	85	사망자수_남자
31	86	사망자수_여자
32	87	PI_P(전산업생산지수-광공업)
33	88	PI_C(전산업생산지수-건설업)
34	89	PI_SI(전산업생산지수-서비스업)
35	90	PI_PA(전산업생산지수-공공행정)
36	91	CD(CD금리91일)
37	92	CP(CP금리91일)
38	93	YT_1(국고채1년)
39	94	YT_3(국고채3년)
40	95	YT_5(국고채5년)
41	96	YT_10(국고채10년)
42	97	CB_3_AA(회사채 3년, AA-)
43	98	CB_3_BBB(회사채 3년, BBB-)
44	99	BaseRate(한국은행 기준금리)
45	100	실업률
46	101	CPI_F(소비자물가지수-식품 및 비주류음료)
47	102	CPI_C(소비자물가지수-의류 및 신발)
48	103	CPI_H(소비자물가지수-가정용품 및 가사서비스)
49	104	CPI_HE(소비자물가지수-보건)
50	105	CPI_T(소비자물가지수-교통)
51	106	CPI_E(소비자물가지수-교육)
52	107	CPI_M(소비자물가지수-기타상품 및 서비스)

□ 분석모델

이번 프로젝트는 머신러닝 모형활용 예측보고서(장나원 외, 2022)의 방법과 유사하게 구간이동(rolling window) 방법을 활용하여 총 16개 모형의 예측성능을 비교하였다. 전체 표본 크기는 112개월(2014년 1월 ~ 2023년 4월)이며, 구간이동 방법을 적용할 때 추정 기간의 크기(window size)는 70개월(5년10개월)으로 설정하고 예측기간은 총 40개월(3년4개월)로 정하였다. 예측 대상기간은 1개월~12개월로 설정하였다.

이번 프로젝트의 분석모델은 머신러닝 모형활용 예측보고서(장나원 외, 2022)에서 활용한 모델인 다음과 같은 총 16개의 시계열 모형과 머신러닝 방법들을 고려하였다.

① RandomWalk, ② 자기회귀(AR Model with Lag 4), ③ LASSO(Least Absolute Shrinkage and Selection Operator), ④ Adaptive LASSO, ⑤ Elastic Net, ⑥ Adaptive Elastic Net, ⑦ Ridge regression, ⑧ CSR(complete subset regression), ⑨ Target Factor, ⑩ Random Forest, ⑪ RF / OLS, ⑫ AdaLasso / RF(Adaptive Lasso를 통해 선택된 설명변수들만을 이용하여 랜덤 포레스트 적용), ⑬ XGBoost, ⑭ NN(Neural Network), ⑮ LSTM(Long-Short Term Memory), ⑯ Boruta_selected_RF(보루타 알고리즘을 통해 중요 설명변수들의 순위를 도출하고 이들 중 교차검증을 통해 정해진 최적 변수의 수만큼만 사용하여 랜덤 포레스트를 적용)의 시계열 및 인공지능의 총 16개 모형의 예측력을 비교분석하였다.³⁾

3) 장나원, 한희준, “머신러닝 방법을 활용한 한국의 수출입 증가율 예측 및 분석”, 국제경제연구 제28권 제4호, 2022, 61~63쪽.

□ 예측력 비교분석

아래의 표는 각 모델의 인구수 예측오차를 나타낸 표이다. 값이 낮을수록 실제값과 유사하게 예측한 것을 의미한다. 즉, 값이 낮을수록 좋은 예측성능을 보이는 모델이다.

<Table1> RMSEs for Population Forecasting

Model	1개월	2개월	3개월	4개월	5개월	6개월	7개월	8개월	9개월	10개월	11개월	12개월
RandomWalk	0.0017	0.0017	0.0015	0.0017	0.0015	0.0015	0.0015	0.0014	0.0016	0.0017	0.0015	0.0015
AR	0.0010	0.0010	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011
Lasso	0.0032	0.0033	0.0032	0.0033	0.0031	0.0026	0.0027	0.0030	0.0029	0.0022	0.0030	0.0026
Adaptive Lasso	0.0030	0.0035	0.0033	0.0030	0.0031	0.0026	0.0027	0.0033	0.0030	0.0020	0.0031	0.0027
Elastic Net	0.0021	0.0027	0.0021	0.0018	0.0027	0.0023	0.0025	0.0026	0.0024	0.0021	0.0027	0.0025
Adaptive Elasticnet	0.0015	0.0019	0.0015	0.0015	0.0026	0.0023	0.0026	0.0029	0.0026	0.0020	0.0024	0.0026
Ridge	0.0012	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0012	0.0012	0.0012	0.0013
CSR	0.0011	0.0012	0.0012	0.0012	0.0012	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013
Target Factor	0.0011	0.0013	0.0013	0.0011	0.0013	0.0013	0.0012	0.0012	0.0013	0.0012	0.0012	0.0015
Random Forest	0.0010	0.0010	0.0010	0.0010	0.0011	0.0011	0.0011	0.0011	0.0012	0.0011	0.0011	0.0012
RF / OLS	0.0012	0.0012	0.0013	0.0013	0.0013	0.0012	0.0011	0.0011	0.0012	0.0011	0.0012	0.0012
AdaLasso / RF	0.0010	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0011	0.0012	0.0012	0.0012	0.0012
XGB	0.0013	0.0014	0.0015	0.0014	0.0012	0.0011	0.0012	0.0013	0.0014	0.0013	0.0014	0.0012
NN	0.0018	0.0018	0.0017	0.0020	0.0022	0.0017	0.0017	0.0016	0.0019	0.0018	0.0015	0.0020
LSTM	0.0016	0.0022	0.0019	0.0019	0.0019	0.0017	0.0020	0.0016	0.0017	0.0019	0.0017	0.0020
Boruta_selected_RF	0.0010	0.0012	0.0012	0.0014	0.0013	0.0012	0.0012	0.0013	0.0013	0.0014	0.0014	0.0011

※ 굵은 글씨는 다른 예측모델들에 비해서 예측오차(RMSE)가 낮은 모델을 의미한다. 모든 개월에서 AR모델이 좋은 예측성능을 보였으며, Random Forest와 AdaLasso/RF 모델도 각각 7개월 예측과 10개월 예측을 제외한 기간에서 우수한 예측성능을 보였다.

머신러닝 모형활용 예측보고서(장나원 외, 2022)에서는 Boruta_selected_RF가 가장 좋은 예측성능을 보였으나, 본 프로젝트에서는 좋은 성능을 발휘하지 못하였다.

▲ 설명변수의 낮은 설명력 추측컨대, 각종 경제지표는 CPI지수에 영향을 미치는 점이 비교적 명확한 반면, 인구수에 영향을 미치는 요인은 워낙 다양하고 불명확한 경우가 대다수여서, 이번 프로젝트에서 설정한 52가지의 변수로는 인구수를 유의미하게 예측하기 어려웠던 것으로 보인다. AR모형은 과거의 인구수 값으로 미래 인구수를 예측하는 모형으로, 이번 프로젝트에서 가장 좋은 성능을 보였다.

□ 모형의 예측력 검증

이제 AR모형의 예측력이 다른 모형들에 비해 통계적으로 우월한지 검토해야한다. 모형예측력 검증에는 Giacomini-White검정과 MCS(Model Confidence Set) 검정을 이용한다.

<Table2> Giacomini-White Test Statistic Results for Population Forecasting

Model	1개월	2개월	3개월	4개월	5개월	6개월	7개월	8개월	9개월	10개월	11개월	12개월
RandomWalk	4.37	4.72	1.74	2.44	3.02	6.06	2.51	6.17	2.99	5.27	6.19	4.72
Boruta_selected_RF	0.69	0.92	0.57	1.04	1.10	0.79	1.58	1.73	0.66	2.48	1.97	0.44
Ridge	1.28	1.16	1.12	1.30	0.32	-0.34	1.08	1.21	1.87	1.33	1.68	2.32
Lasso	5.00	4.89	5.90	5.94	6.95	4.51	3.36	3.50	6.82	4.83	3.70	3.36
Adaptive Lasso	4.43	4.77	5.31	6.14	5.65	4.72	3.99	3.21	5.47	4.99	3.30	3.17
Elasticnet	3.63	4.41	1.36	2.21	2.76	3.70	3.11	3.30	3.64	4.55	4.65	3.39
Adaptive Elasticnet	2.33	4.21	0.81	1.60	2.24	3.95	3.65	3.08	3.41	5.21	5.05	3.21
CSR	1.90	1.76	0.33	1.54	1.14	0.32	1.30	2.99	1.02	1.24	1.93	2.16
Random Forest	0.97	0.03	-0.81	-1.03	-0.23	-1.41	0.99	0.40	1.61	-0.19	0.92	1.46
XGBoost	1.97	2.83	2.26	1.01	2.25	-0.86	2.28	2.16	2.87	2.13	2.05	1.46
NN	4.54	3.70	3.45	3.90	4.45	1.83	2.78	5.69	2.72	3.54	3.62	3.44
LSTM	3.49	4.28	2.25	5.27	3.92	3.87	5.67	3.11	2.36	6.69	3.81	3.43

이 표는 예측값의 절대오차 기준으로 AR모형과 각각의 다른 모형들의 Giacomini-White Test 통계량 표이다. 양의 값은 비교대상 모형들의 예측오차가 AR모형의 예측오차보다 크다는 것을 의미한다. Random Forest의 3~6개월 예측오차보다 성능이 안좋은 경우도 발견되었지만, 대부분의 경우 AR모형의 예측오차가 더 작은 것을 볼 수 있다.

<Table3> Giacomini-White Test p-value Results for Population Forecasting

Model	1개월	2개월	3개월	4개월	5개월	6개월	7개월	8개월	9개월	10개월	11개월	12개월
RandomWalk	0.00	0.00	0.08	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Boruta_selected_RF	0.49	0.36	0.57	0.30	0.27	0.43	0.11	0.08	0.51	0.01	0.05	0.66
Ridge	0.20	0.25	0.26	0.19	0.75	0.73	0.28	0.23	0.06	0.18	0.09	0.02
Lasso	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Adaptive Lasso	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Elasticnet	0.00	0.00	0.17	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Adaptive Elasticnet	0.02	0.00	0.42	0.11	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CSR	0.06	0.08	0.74	0.12	0.25	0.75	0.19	0.00	0.31	0.21	0.05	0.03
Random Forest	0.33	0.98	0.42	0.30	0.82	0.16	0.32	0.69	0.11	0.85	0.36	0.14
XGBoost	0.05	0.00	0.02	0.31	0.02	0.39	0.02	0.03	0.00	0.03	0.04	0.14
NN	0.00	0.00	0.00	0.00	0.00	0.07	0.01	0.00	0.01	0.00	0.00	0.00
LSTM	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00

이 표는 예측값의 절대오차 기준으로 AR모델의 예측성능이 비교대상 모델의 예측성능보다 유의수준 5% 수준에서 통계적으로 유의미하게 좋은지의 여부를 나타낸 것이다. 색깔이 없는 칸은 AR모델과 비교대상 모델의 예측성능이 유의수준 5% 기준 통계적으로 유의미한 차이를 가지는 경우이다. 색깔이 표시된 칸은 AR모델과 비교대상 모델의 예측성능이 통계적으로 유의미한 차이를 가지지 않는 경우이다. Ridge모델과 RandomForest모델을 제외한 다른 모델과의 비교에서 대부분의 기간예측에 대해 AR모델의 예측성능이 우수한 것으로 나타났다.

<Table4> MCS(Model Confidence Set) Test Results for Population Forecasting

Model	1개월	2개월	3개월	4개월	5개월	6개월	7개월	8개월	9개월	10개월	11개월	12개월
RandomWalk												
AR	○	○			○	○	○	○	○	○	○	○
Ridge												
Lasso												
Adaptive Lasso												
Elasticnet												
Adaptive Elasticnet												
CSR												
Random Forest			○	○		○				○	○	○
XGBoost												
Boruta_selected_RF												
NN												
LSTM												

이 표는 유의수준 50% 수준에서 MCS Test 결과를 보여준다. 만약 모델이 MCS 표에 포함되어있다면, 그 모델이 가장 좋은 예측성능을 보인다는 의미이다. AR(4) 모델이 3,4개월만을 제외한 모든 예측 대상기간에서 AR(4)모델이 포함되어 있어, 16개의 시계열 예측모델 중 비교적 예측성능이 우수한 모델임을 알 수 있다.

Ⅲ. 인구예측 결과

□ 향후 인구예측

이제 이번 프로젝트에서 가장 중요한 실제 인구예측을 진행해보고자 한다. 위 과정에서 AR(4) 모델이 비교적 좋은 성능을 보인 것으로 확인하였다. AR(4)모델로 1~12개월의 향후 인구를 예측해보고자한다. 화성시 인구수 추이는 우상향 추세를 가진 비정상데이터이기에, 바로 예측하는 것은 통계적으로 적절하지 않다. 정상성데이터로 변환한 인구성장률을 예측하여, 이를 바탕으로 인구추이를 예측하는 것이 적절하다.

<Table5> 화성시 인구추이 예측

순번	연월	예측 인구성장률	총인구수(명)
1	2023년 4월	—	978,653
2	2023년 5월	0.0038899	982,460
3	2023년 6월	0.0041965	986,583
4	2023년 7월	0.0048495	991,367
5	2023년 8월	0.0044766	995,805
6	2023년 9월	0.0033211	999,112
7	2023년 10월	0.0028119	1,001,922
8	2023년 11월	0.0027496	1,004,677
9	2023년 12월	0.0023012	1,006,989
10	2024년 1월	0.0025762	1,009,583
11	2024년 2월	0.0026093	1,012,217
12	2024년 3월	0.0022981	1,014,543
13	2024년 4월	0.0025841	1,017,165

예측결과, 화성시 인구 100만 도달시기는 2023년 9월과 10월 사이가 될 것으로 예상된다. 이상으로 다양한 시계열분석모델과 인공지능 모델을 활용하여 인구추이를 예측분석해보았다. 시계열예측은 지금까지의 데이터로 미래 데이터를 예측하기에, 지금까지 없었던 변수가 생길 경우 시계열흐름은 요동칠 수 있다. 예를 들어, 타 지자체 거주 화성시 공무원들에게 화성시로의 전입 시 인센티브를 주는 정책을 펼치거나, 타 지자체보다 급격히 아동양육 지원금액을 대폭 향상시키거나, 외국인 근로자 채용 지원금 확대 정책시행 등 지금까지 없었던 변수가 발생하게 되면, 이에따라 시계열 예측결과는 다소 빗겨갈 수 있음을 인지해야한다.

□ 개발사업 · 주택건설사업 현황

택지개발 및 주택건설에 진행속도에 따라 100만 도달시기가 달라질 수 있다. 100만 도달 예측시기까지(23년 10월) 준공되는 개발사업은 없었으며, 주택건설사업의 경우 4개건이 준공 예정이며 신규유입인구는 약 6,200명으로 예상된다. 이는 올해 10월에 100만 인구 도달에 긍정적인 신호로 판단된다.

<자료 : 2023년도 1분기 화성시 시정현황>

○ 택지개발사업

(단위: 천㎡, 세대, 명)

사 업 명	면적	세대수	계획인구	일정별 추진계획					사 업 시 행 자
				지구지정	개발계획	실시계획	착 공	준 공	
계	3,576	22,592	55,489	-					
동탄지구 (2단계)	812	489	1,223	*01.04.30.	*01.12.14.	*02.12.26.	*03.03.31.	(*24.12.)	LH
동탄(2)지구 (4단계)	1,572	18,340	42,038	*07.12.20.	*08.07.11.	*10.07.07.	*11.04.18.	(*24.12.)	LH GH
태안3지구 (1단계)	595	2,961	9,480	*98.05.11.	*03.05.06.	*05.01.10.	*06.02.28.	(*23.04.)	LH
태안3지구 (2단계)	597	802	2,748	*98.05.11.	*03.05.06.	*05.01.10.	*06.02.28.	(*23.12.)	LH

※ 사업완료

발안지구(204천㎡/6,380명), 태안지구(1,168천㎡/35,256명), 봉담1지구(774천㎡/16,941명), 동탄지구(1단계:8,223천㎡/124,326명), 향남1지구(1,698천㎡/31,531명), 향남2지구(3,152천㎡/46,231명) 동탄(2)지구(1단계:1,703천㎡/65,660명 2단계:3,089천㎡/116,879명 3단계:1,473천㎡/61,301명)

○ 공공주택지구조성사업

(단위: 천㎡, 세대, 명)

사 업 명	면적	세대수	계획인구	일정별 추진계획					사 업 시 행 자
				지구지정	개발계획	지구계획	착 공	준 공 (예정)	
계	8,418	58,614	140,138	-					
비봉지구 (1단계)	820	7,090	16,641	*07.04.30.	*13.08.23.	*14.09.23.	*16.09.26.	(*22.12.)	LH
비봉지구 (2단계)	43	-	-	*07.04.30.	*13.08.23.	*14.09.23.	*16.09.26.	(*24.06.)	LH
어천지구	744	4,138	9,770	*18.12.31.	*20.12.30.	*20.12.30.	(*24.08.)	(*26.06.)	LH
봉담3지구	2,286	17,526	42,063	*22.12..20.	(*24.06.)	(*24.06.)	(*25.09.)	(*30.06.)	LH
진안지구	4,525	29,860	71,664	(*23.10.)	(*24.06.)	(*24.06.)	(*25.09.)	(*30.06.)	LH

※ 사업완료 : 봉담2지구(1,438천㎡/28,716명)

○ 도시개발사업

(단위: 천㎡, 세대, 명)

사업명	면적	세대수	계획인구	일정별 추진계획					사업시행자
				지구지정	개발계획	실시계획	착공	준공	
계	1,119	10,117	25,681	-					
병점복합타운	376	1,588	3,194	*05.12.15.	*07.12.13.	*10.04.05.	*16.03.01.	(*24.12.)	LH
반월지구	138	1,306	3,267	*17.08.03.	*18.09.19.	*19.08.20.	*20.04.13.	(*23.02.)	반월지구 도시개발 사업조합
동화지구	299	2,447	6,801	*15.10.23.	*15.10.23.	*18.12.27.	*20.04.20.	(*23.06.)	동화지구 도시개발 사업조합
도이지구	40	547	1,423	*19.11.11	*19.11.11	(*22.10.)	(*23.03.)	(*28.01.)	향남 도이지구 도시개발 사업조합
내리지구	266	4,229	10,996	*19.12.17.	*19.12.17.	*20.10.16.	*21.04.09.	(*24.07.)	봉담 내리지구 도시개발 사업조합

※ 사업완료 : 남양지구(682천㎡/4,929명), 조암지구(245천㎡/3,465명), 남양뉴타운(2,573천㎡/39,111명)

○ 복합도시개발사업

(단위: 천㎡, 세대, 명)

사업명	면적	세대수	계획인구	일정별 추진계획					사업시행자
				지구지정	개발계획	실시계획	착공	준공	
계	55,644	60,000	150,000	-					
송산그린시티	55,644	60,000	150,000	*98.11.14.	*08.03.14.	*11.09.27.	*11.11.20.	(*30.12.)	수자원 공사

○ 주택건설사업 현황

순번	행정구역	사업명칭	세대수	계획인구	사용검사 예정일
1	봉담(2)	중흥S클래스아파트(B-4)	824	2,472	2023년 8월
2	봉담(2)	중흥S클래스아파트(B-1)	1,050	3,150	2023년 8월
3	송산그린시티	송산 리안비채(EB4)	128	282	2023년 9월
4	송산그린시티	송산 리안비채(EB5)	131	288	2023년 9월
5	동탄(2)	힐스테이트 동탄 더 테라스	125	275	2023년 11월
6	봉담(2)	중흥S클래스아파트(C-1)	707	2,121	2023년 12월
7	봉담(2)	중흥S클래스아파트	808	2,424	2023년 12월
8	송산그린시티	메종 드 엘리프 송산(EB2)	116	255	2023년 12월
9	송산그린시티	메종 드 엘리프 송산(EB3)	88	194	2023년 12월
10	동탄(2)	자연앤 이편한세상	1,227	3,681	2023년 12월
11	남양뉴타운	화성남양뉴타운 B16블럭 공동주택	600	1,800	2023년 12월

IV. 참고문헌

- 김형기, 문경중, “한국의 시도별 장래인구 예측”, 『국토계획』 제46권 제6호(2011).
- 장나원 · 한희준, “머신러닝 방법을 활용한 한국의 수출입 증가율 예측 및 분석”, 『국제경제연구』 제28권 제4호, 2022.
- Medeiros, M. C., Vasconcelos, G. F., Veiga, A. , and Zilberman, E.(2019), “Forecasting Inflation in a Data-rich Environment: The Benefits of Machine Learning Methods” , Journal of Business & Economic Statistics, 1-45.
- Marco Peixeiro(2022), “Time Series Forecasting in Python” , Manning Publications.

분석코드

- Medeiros et al. (2019) (<https://github.com/gabrielrvsc/>)
- Kim and Han(2022) (<https://github.com/bryank2242/Forecasting-VIX>)