06

경기도 화성시

화성특례시 도약을 위한 인구 시계열예측

06

화성특례시 도약을 위한 인구 시계열예측

경기도 화성시

기획조정실 스마트도시과 빅데이터팀 이동재

요 약

화성시는 2023년 말에서 2024년 초 사이에 인구 100만 명의 대도시에 진입할 것으로 예상된다. 연말 기준 인구수가 2년 연속 100만 명 이상인 경우 특례시로 지정될 수 있기 때문에 특례시 지정을 위한 행정절차 등을 진행하기 위해 화성시의 100만 인구 도달 시기를 정확하게 예측하는 것이 필요하였다.

인구학적 요소와 인구 유입을 고려하여 화성시의 미래 인구 추이를 예측하고자 거시환경, 경제환경 등 다양한 변수를 적용한 예측모델을 도출하고, 이 모델을 활용한 데이터 분석으로 100만 인구 도달 시기를 보다 정확하게 예측하려고 하였다.

이 분석에서는 전통적인 통계모델과 인공지능 시계열예측모델을 포함한 16개 인구 예측모델의 성능을 비교 분석하였다. 화성시 인구 데이터와 그에 영향을 미치는 거시환경 변수, 택지개발사업, 도시개발사업, 공공주택지구 조성사업 등의 데이터를 활용하여 데이터 수집, 전처리, 탐색, 모델링, 검증 및 평가, 시계열예측 및 시각화 순으로 진행하였다.

16개 모델을 통한 분석 결과, 4차자기회귀모델이 가장 좋은 성능을 보였다. 인구에 영향을 미치는 변수를 도출하는 것보다 자신의 과거 데이터를 독립변수로 설정하는 자기회귀모델이 가장 높은 성능을 발휘한 것이다.

이 모델을 이용해 향후 1년간의 인구를 예측한 결과, 화성시 인구가 100만에 도달하는 시기는 2023년 10월로 예상되었다. 분석 결과는 행정절차의 기준 시점을 제시하고 행정에서의 불확실성을 줄여 관련 행사 등 다양한 정책 기획에 활용할 계획이다.

가 | 분석 개요

추진 배경

- 인구가 100만 명에 도달하는 시기에 대한 예측 범위가 넓음
  - 특례시 인정 기준인 100만 인구 달성 시기 예측 범위가 올해 말~내년 초로 광범위하여 구체적인 시기 예상이 어려웠음
  - 특례시 관련 행정절차를 진행하기 위해 정확한 기준 시기를 도출할 필요성 대두

분석 필요성 및 목적

- 다양한 인구학적 요인과 인구 유입을 고려하여 화성시 향후 인구 추이 예측
  - 거시환경, 경제환경 등 다양한 변수를 고려한 예측모델 개발
  - 데이터 기반 예측을 통해 지역 특화 정책 방향 제시

기대 효과

- 저출산 및 고령화에 따른 급격한 인구변화의 사회적 영향 예측
  - 100만 인구 도달 시기의 예측 정밀화

    ▶ (기존) 2023년 말~2024년 초 → 2023년 10월

나 | 분석 설계

요구 사항

- 인구 예측모델 중 가장 좋은 성능의 모델 도출 및 예측
  - 전통적 통계모형 및 인공지능 시계열예측모델 사용
    ▶ 16개 모델의 성능을 비교 분석

분석 목록

- 거시환경, 경제상황 등 인구에 영향을 주는 변수 수집
  - 토지가격, 아파트 매매가격, 각종 금리지수 등 106개 변수 데이터 확보
  - 각 변수별 영향도 측정

분석 데이터

- 분석 대상
  - 화성시 인구 데이터와 그에 영향을 주는 거시환경 변수
    ▶ 택지개발사업, 도시개발사업, 공공주택지구조성사업 반영
- 분석 범위
  - 공간적 범위: 경기도 화성시
  - 시간적 범위: 2014.1.~2023.5.

• 분석 활용 데이터

데이터명	형태	내용	출처	기준연도	내·외부 데이터
개발사업 계획	hwp	택지개발사업, 공공주택 조성사업, 도시개발사업	신도시재생과, 지역개발과	2023	내부
인구 데이터	csv	내·외국인, 전입, 전출, 출생, 사망	행정안전부, 법무부	2014 ~2023	외부 (공공)
거시환경 데이터	csv	지가변동률, 아파트가 변동률 (매매, 전세, 월세)	한국부동산원	2014 ~2023	외부 (공공)
경제상황 데이터	csv	전산업생산지수, CD금리, CP금리, 국고채, 회사채, 기준금리, 실업률, 소비자물가지수	한국은행	2014 ~2023	외부 (공공)

〈 표 6-1 〉 활용 데이터 목록

전체 프로세스(도식화)

〈 그림 6-1 〉 분석 프로세스

세부 프로세스(분석 방법론)

• 데이터 수집

- 내국인 인구 데이터(행정안전부), 외국인 인구 데이터(법무부), 아파트가 데이터(한국부동산원), 경제 데이터(한국은행 통계정보시스템) 및 담당 부서를 통한 데이터 수집

  ▶ 106종의 데이터 확보

• 데이터 전처리

- 결측치 탐색 및 결측치 처리

  ▶ 포워드 필링(Forward Filling) 방식 사용

• 데이터 탐색 및 분석 모델링

- 정상성 여부 탐색(단위근 검정), 정상성 데이터로의 변환(단위근 제거)

▶ 단위근 검정 방법: ADF Test, KPSS Test

▶ 단위근 제거: 차분 및 로그 변환

• 데이터 분석 및 시각화

- 16개 모델로 시계열예측 진행 및 성능 비교

① RandomWalk ② 자기회귀(AR Model with Lag 4) 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

〈 표 6-2 〉 시계열예측모델

다 | 분석 결과

모델 성능 비교

- 모든 개월에서 자기회귀(AR) 모델이 좋은 예측성능을 보임. 랜덤 포리스트(Random Forest)와 AdaLasso/RF 모델도 각각 7개월 예측과 10개월 예측을 제외한 기간에서 우수한 예측성능을 보임

※ 은 다른 예측모델에 비해 예측오차(평균제곱근오차, RMSE)가 낮은 경우임

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 Elastic Net	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	0.002
LSTM	0.0016	0.0022	0.0019	0.0019	0.0019	0.0017	0.0020	0.0016	0.0017	0.0019	0.0017	0.002
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

〈 표 6-3 〉 인구 예측모델별 예측오차(평균제곱근오차, RMSE) 비교

- 예측값의 절대오차 기준으로 자기회귀(AR) 모델의 예측성능이 비교대상 모델의 예측성능보다 유의수준 5% 수준에서 통계적으로 유의미하게 좋은지의 여부를 나타냄
  - 색이 없는 칸은 자기회귀(AR) 모델과 비교대상 모델의 예측성능이 유의수준 5% 기준 통계적으로 유의미한 차이를 가지는 경우임
  - 리지(Ridge) 모델과 랜덤 포리스트(Random Forest) 모델을 제외한 다른 모델과의 비교에서 대부분의 기간 예측에 대해 자기회귀(AR) 모델의 예측성능이 우수한 것으로 나타남

※ 은 자기회귀(AR) 모델과 비교대상 모델의 예측성능이 통계적으로 유의미한 차이가 없는 경우임

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	0.2	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
Elastic Net	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 Elastic Net	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

〈 표 6-4 〉 인구 예측모델별 자코미니-화이트 검정(Giacomini-White Test) p 결과값 비교

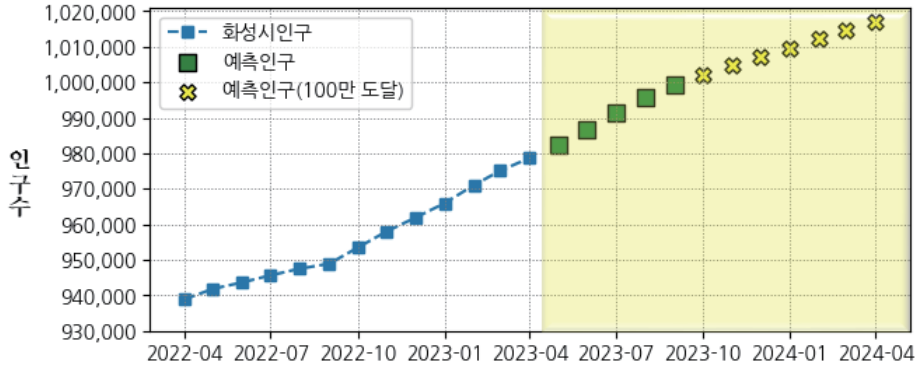
분석 결과 요약(시각화)

- 2014년 1월~2023년 4월까지의 주어진 데이터로 미래의 인구를 예측
- 16개 모델 평가 결과 4차자기회귀(AR(4)) 모델이 비교적 좋은 성능을 보임
  - 4차자기회귀(AR(4)) 모델로 1개월~12개월의 향후 인구를 예측
  - 화성시 인구수 추이는 우상향 추세를 가진 비정상 데이터이기에, 바로 예측하는 것은 통계적으로 적절하지 않음. 정상성 데이터로 변환한 인구성장률을 예측하여 인구 추이를 예측하는 것이 적절

• 예측 결과, 화성시 인구 100만 도달 시기는 2023년 10월로 예상

연번	연월	예측 인구성장률	인구수(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

〈 표 6-5 〉 화성시 인구 추이 예측



〈 그림 6-2 〉 화성시 인구 예측 그래프

라 | 시사점 및 한계점

시사점

- 행정의 가장 기본 데이터인 인구 데이터를 통해 미래 예측 역량 강화
  - 막연하고 불확실한 미래에 대한 예측 기술 확보
  - 정책에서의 불확실성 제거 및 기준점 마련

한계점

- 인구에 영향을 미치는 변수를 도출하는 것이 현실적으로 어려움
  - 인구에 영향을 미치는 변수는 매우 다양하고 복합적으로 작용

※ 독립변수를 자신의 과거 데이터로 설정하는 자기회귀모델을 활용하는 것이 가장 높은 성능을 발휘

마 | 정책 활용 및 향후 계획

행정 적용 사항

- 100만 도시 준비를 위한 각종 행정절차의 기준 시점 제시
  - 데이터 기반 미래 예측을 통한 행정 추진 도모
  - 행사 시기 조정 등 업무 계획 조정

정책 활용 방안

- 향후 인구, 부동산, 각종 경제지수 등 예측을 통한 정책 기획
  - 데이터별 모델 성능 비교 후 높은 성능 모델 선택·활용

향후 계획

- 화성시 영유아 정책 평가 및 영유아 수 예측 보고서 발간
  - 영유아 정책 예산이 영유아 수에 미치는 영향에 대한 연구 추진
  - 보육 정책 평가 전문화 및 보육 정책 계획