접수번호	※작성하지 않음

출산율에 영향을 주는 요인분석과 출산율 증가 방향성 제안

제 목	출산율에 영향을 주는 요인분석과 출산율 증가 방향성 제안

신 청 자 명	소속/직위	무소속	성 명	박주안, 박거량, 송재근
	휴대전화	010-3743-7614	전자우편	johnnyworld@naver.com
제출일	2023.07.21.			

출산율에 영향을 주는 요인분석과 출산율 증가 방향성 제안

1. 주제선정 및 연구배경

“대한민국의 저출산 문제가 개선되지 않는다면 현재 지구상에서 소멸하는 첫 국가가 될것이다.” 라는 위기감을 일깨워준 인구학자 데이비드 콜먼(David Coleman)은 다음과 같이 주장했다. “인구감소는 전 세계적인 추세이지만 대한민국의 출산율 감소를 보면 2070년에 국가 소멸이라는 절대적 위기에 처할 수 있다.” [1] 대한민국은 1983년 이후 거의 40년간 저출산 현상이 지속되고 있다. 통계청의 2023 인구동향조사에 따르면, 올해 1분기 합계출산율은 0.81명으로 기록되었다.[2] 경제협력개발기구(OECD) 평균인 1점대를 한참 밑도는 기록이다.[3]

정부는 인구감소에 따른 변화에 대응하기 위해 2006년부터 「저출산・고령사회 기본법」 5년 주기로 정책을 시행하고 있다.[4] 특히 2021년부터 시행되고 있는 “제4차 「저출산고령사회기본계획」” 은 기존에 저출산 대책을 위한 예산 투자 뿐만 아니라, 근본적인 문제 해결을 제시하고 있다. 개인 삶의 질 개선, 남녀 성 평등, 저출산과 고령화에 대한 인구변화 대응을 목표로 삼고 출산 장려 정책을 추진하고 있다.[5] 특히 청년층이 기회를 잡을 수 있도록 보장하는 자립 기반 정책이 강화되었다. 이를테면, 청년 맞춤 임대주택 24만개 공급, 청년 임차가구 40만 가구 금융지원, 경력단절 예방 서비스 확대, 청년 진로설계 상담등 주택, 금융지원, 경제활동지원을 통해 출산율 향상에 지원하고 있다.

본연구는 출산율에 영향을 줄 수 있는 다양한 요인들을 분석하고, 현재 진행되고 있는 저출산에 대한 정부정책의 유의성에 대해 파악하였다. 구체적으로 건강의료, 경제활동, 교육, 주택, 혼인 및 출생과 관련된 데이터를 기반으로 분석하였다. 출산율에 가장 유의미한 요인들을 기반으로 출산율 대책 마련에 도움 될 수 있는 기대효과와 방향성을 제안하겠다.

1) 양봉모, “인구학자 데이비드 콜먼 교수 "한국 저출산 지속되면 2070년 국가 소멸 위험”, BBSNews, 2023.5.17.
2) 통계청(2023), 「2023년 3월 인구동향」.
3) OECD (2022), Fertility rates. accessed 7월 10일 검색.
4) 김우림 (2021), 저출산 대응 사업 분석·평가. 서울: 국회예산정책처.
5) 보건복지부 인구정책총괄과 (2020), 제4차 저출산·고령사회 기본계획.

2. 데이터 분석

2.1. 활용데이터

출산율에 영향을 주는 요인분석에 활용된 데이터는 다음과 같다.

데이터명	출처	비고
합계출산율(시도/시/군/구)	통계청 인구동향조사	인구동향조사
시도/시군구/월별 혼인	통계청 인구동향조사	인구동향조사
인구 천명당 의료기관 종사 의사수(시도/시/군/구)	건강보험심사평가원 자원평가실	e-지방지표
건강생활실천율(시도/시/군/구)	질병관리청 지역사회건강조사	e-지방지표
스트레스 인지율(시도/시/군/구)	질병관리청 만성질환관리국 만성질환관리과	e-지방지표
유치원수(시도/시/군/구)	한국교육개발원	e-지방지표
초등학교수(시도/시/군/구)	한국교육개발원	e-지방지표
경제활동인구(시/군/구)	통계청 지역별고용조사	e-지방지표
경제활동인구(시/도)	통계청 경제활동인구조사	e-지방지표
고용률(시/군/구)	통계청 지역별고용조사	e-지방지표
고용률(시도)	통계청 경제활동인구조사	e-지방지표
청년고용률(시/군/구)	통계청 지역별고용조사	e-지방지표
아파트매매가격지수(시도/시/군/구)	한국부동산원 부동산통계처 주택통계부	e-지방지표
아파트전세가격지수(시도/시/군/구)	한국부동산원 부동산통계처 주택통계부	e-지방지표
주택매매가격변동률(시도/시/군/구)	한국부동산원,	e-지방지표
주택매매가격지수(시도/시/군/구)	한국부동산원 부동산통계처 주택통계부	e-지방지표
주택전세가격지수(시도/시/군/구)	한국부동산원 부동산통계처 주택통계부	e-지방지표

표1. 분석에 사용된 데이터

모든 데이터는 통계청 인구동향조사와 e-지방지표를 활용하였다. 합계출산율을 종속변수(y)로 설정하고 나머지 데이터를 독립변수(x)로 지정하였다.

2.2. 데이터 전처리

전처리 과정은 파이썬(Python)언어와 판다스(Pandas) 라이브러리를 활용해 진행하였다. 총 17개의 데이터를 확인해 본 결과 지수로 나타낸 데이터와 비율로 나타낸 경우가 있었고, 그 외에도 다양한 문제점들이 발견되었다. 먼저 데이터의 시점을 확인했을 때 다양한 형식들이 있었다. 데이터마다 연도, 월, 분기까지 다양하게 표시된 데이터들을 확인하였고, 일관성 있는 데이터 통합을 위해 시점을 모두 연도로 통일하였다(그림1).

```
k = 0
m = 0
r = pd.DataFrame(columns=['시점','행정구역별','청년고용률'])
for i in range(2013,2023):
    l = len(q[q['시점'] == f'{i}.1/2']['행정구역별'])
    for j in range(0,l):
        if(q[q['시점'] == f'{i}.1/2']['행정구역별'][k] != q[q['시점'] == f'{i}.2/2']['행정구역별'][k+l]):
            print('error')
        else:
            r.loc[m] = [i,q[q['시점'] == f'{i}.1/2']['행정구역별'][k], (q[q['시점'] == f'{i}.1/2']['청년고용률'][k]+q[q['시점'] == f'{i}.2/2']['청년고용률'][k+l])/2]
        k += 1
        m += 1
    k += 1
```

그림1. 시점 데이터 통합 전처리 과정

다음으로, 빈 데이터 값(NaN)가 있는 경우 제외하였다. 행정구역의 통합과 지명 변경의 문제가 있다. 2010년에 마산, 창원, 진해시가 창원시로 통합된 사례가 있어, 모두 창원시로 변경해 주었다(그림2).

```
df = pd.read_csv(filepath[5], encoding = 'ANSI')
df['행정구역별'] = df.apply(lambda x: np.NaN if x['행정구역별'] == '마산시' else x['행정구역별'] , axis = 1)
df['행정구역별'] = df.apply(lambda x: np.NaN if x['행정구역별'] == '창원시' else x['행정구역별'] , axis = 1)
df['행정구역별'] = df.apply(lambda x: np.NaN if x['행정구역별'] == '진해시' else x['행정구역별'] , axis = 1)
df['혼인신고수'] = df.apply(lambda x: np.NaN if x['혼인신고수'] == '-' else x['혼인신고수'] , axis = 1)
df = df.dropna(axis=0,how='any')
df.reset_index(drop = True, inplace = True)
df.to_csv('C:/Users/user/Downloads/시도_시군구_월별_혼인.csv', index = False, encoding = 'ANSI')
```

그림2. 행정구역별 데이터 통합 전처리 과정

행정구역명의 통합성을 위해 제주특별자치도와 강원특별자치도는 제주도와 강원도로 변경하였다(그림3).

```
before = ['제주특별자치도','강원특별자치도']
after = ['제주도','강원도']
for i in range(0,len(filepath)):
    df = pd.read_csv(f"{filepath[i]}",encoding='ANSI')
    for j in range(0,2):
        df['행정구역별'] = df.apply(lambda x: after[j] if x['행정구역별'] == before[j] else x['행정구역별'] , axis = 1)
    df.to_csv(f"C:/Users/user/Downloads/pre3/{filepath[i].split('pre2')[-1][1:]}",index=False,encoding='ANSI')
```

그림3. 행정구역명 데이터 통합 전처리 과정

독립변수(x)에 해당하는 데이터 중 “경제활동인구_계”, “유치원수” 등은 해당 행정구역의 인구수가 고려되지 않고 단순 크기로 보여준 사례가 있었다. 실제 인구수를 고려해야 다른 변수들과 비교하기 수월할 거라 판단하였다. 이를 통해, 해당 행정구역의 인구수로 나눈 비율로 변경하여 사용하였다(그림4).

```
new_full_data['경제활동인구_계_비율'] = new_full_data['경제활동인구_계'] / new_full_data['계 (명)']
new_full_data['경제활동인구_남자_비율'] = new_full_data['경제활동인구_남자'] / new_full_data['남 (명)']
new_full_data['경제활동인구_여자_비율'] = new_full_data['경제활동인구_여자'] / new_full_data['여 (명)']
new_full_data['혼인신고수_비율'] = new_full_data['혼인신고수'] / new_full_data['계 (명)']
new_full_data['출생아수_비율'] = new_full_data['출생아수'] / new_full_data['계 (명)']
new_full_data['유치원수_비율'] = new_full_data['유치원수'] / new_full_data['계 (명)']
new_full_data['의료기관_종사_의사수_비율'] = new_full_data['의료기관_종사_의사수'] / new_full_data['계 (명)']
new_full_data['초등학교수_비율'] = new_full_data['초등학교수'] / new_full_data['계 (명)']
new_full_data
```

그림4. 지수를 비율로 변환한 전처리 과정

2.3. 통계분석 및 시각화

전처리한 데이터를 통계분석에 앞서 저출산에 대한 사회적 시각을 파악하였다. 웹크롤링(Web Crawling)과 워드 클라우드(Word Cloud)를 통해 확인하였다. 형태소 분석을 위해 Konlpy를 사용하였고, Seaborn과 wordcloud 라이브러리를 활용해 시각화 하였다. 네이버 API 중 기사 검색을 활용하였고, 출산율과 관련된 기사 제목의 단어 빈도수만 집계한 결과는 다음과 같다(그림5).

그림5. 출산율 기사 제목 기반 워드클라우드 시각화 결과

데이터분석은 3가지 단계로 진행하였다. 선형회귀(Linear regression), 다중선형회귀(Multiple Linear regression), 다중공선성(Multicollinearity)을 파악하기 위한 VIF(Variance Inflation Factors)와 주성분 분석(Principal Component Analysis, PCA)이 있다.

먼저, 선형회귀(Linear regression)를 사용하여 특정 컬럼(x)이 합계출산율(y)에 얼마나 영향을 미치는지 기울기 값(coef_옵션)을 통해 확인하였다. 기울기를 파악할 때 y값은 모두 합계출산율로 동일하다. 반면 x값은 서로 다른 수치를 가지고 있어 일정 범위 안의 값으로 설정해줄 필요가 있다. 이때 Scaler를 활용해야 하며, 보편적으로 사용되는 Standard Scaler를 활용하였다. Standard Scaler는 평균이 0이고 표준편차가 1인 값으로 데이터의 값을 조정(scaling)한다.

스케일링 이후 학습에 들어간다. 여기서 Linear regression 모델의 경우, x값은 한 컬럼에 대해 스케일링 된 값이고, y값은 스케일링하지 않은 합계출산율을 사용한다. 기울기를 구하면 합계출산율과 얼마나 밀접한 연관이 있는지 파악할 수 있다. 이를테면 혼인신고 수 비율과 출생아 수 비율의 기울기는 0.147과 0.176이다. 한편 혼인신고 수 비율과 출생아 수 비율은 합계출산율에 대한 지표로 쓰이기에는 합계출산율과 비슷한 의미를 지니는 직접적인 컬럼이기에 배제하였다. 혼인신고 수 비율, 출생아 수 비율의 기울기와 비교했을 때 유의미한 기울기값을 갖는 컬럼이라고 판단하였다.

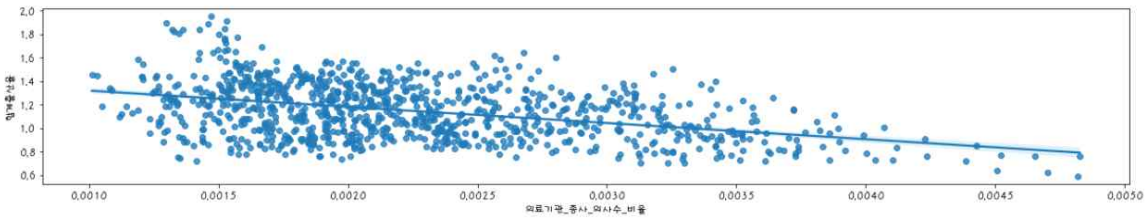

그림6. 의료기관 종사 의사수 비율(-0.095)

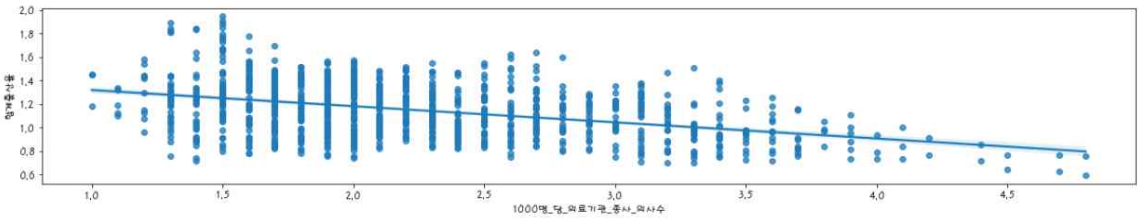

그림7. 1000명당 의료기관 종사 의사 수(-0.094)

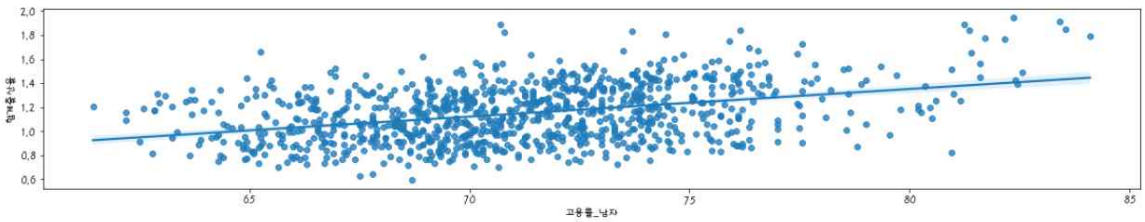

그림8. 고용률 남자(0.088)

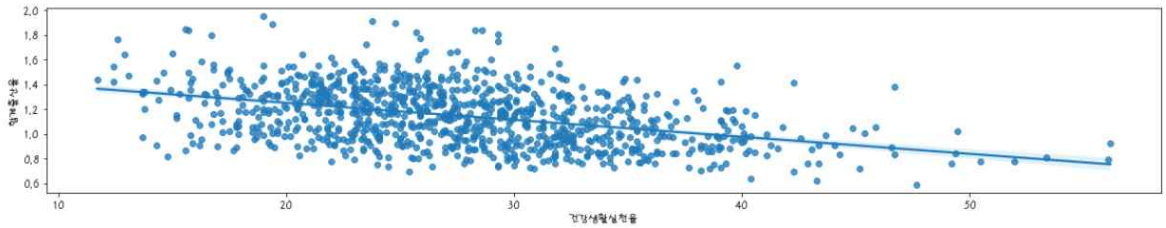

그림9. 건강생활실천율(-0.094)

선형회귀를 통해 유의미한 컬럼들(x)을 확인할 수 있었다. 하지만 각각의 컬럼과 합계출산율의 개인적인 관계만을 그린 그래프이기에 다른 컬럼들이 미치는 영향을 고려되지 않았다. 다중선형회귀는 일반적으로 독립변수(x)가 여러 개일 때 종속변수(y)의 기울기 변화를 확인할 수 있기에 적합하다고 판단하여 분석해 보았다. x값에 해당하는 컬럼들 중 경제활동인구_계, 유치원수 등과 같은 컬럼들은 해당 행정구역의 인구수가 고려되지 않은 단순 크기를 나타냈다. 일률적인 데이터 활용을 위해 해당 행정구역의 인구수로 나눈 비율로 변경하여 사용하였다. y값은 선형회귀와 동일하게 스케일링(scaling)하지 않은 합계출산율을 사용하였다. 각 컬럼의 가중치 값을 토대로 컬럼이 합계출산율에 미치는 영향을 파악하였고 다음 데이터프레임과 그래프를 통해 확인하였다(그림10).

	가중치	가중치(절댓값)
스트레스_인지율	-0.001177	0.001177
건강생활실천율	-0.018140	0.018140
의료기관_종사_의사수_비율	-0.019553	0.019553
유치원수_비율	-0.024061	0.024061
청년고용률	-0.032320	0.032320
혼인신고수_비율	-0.055207	0.055207
초등학교수_비율	0.080809	0.080809
고용률_남자	0.124312	0.124312
주택매매가격지수	-0.159952	0.159952
아파트전세가격지수	-0.173790	0.173790
고용률_여자	0.175643	0.175643
주택전세가격지수	0.178241	0.178241
아파트매매가격지수	0.180895	0.180895
고용률_계	-0.232540	0.232540
출생아수_비율	0.243926	0.243926
경제활동인구_남자_비율	-0.265388	0.265388
경제활동인구_여자_비율	-0.278636	0.278636
경제활동인구_계_비율	0.475802	0.475802

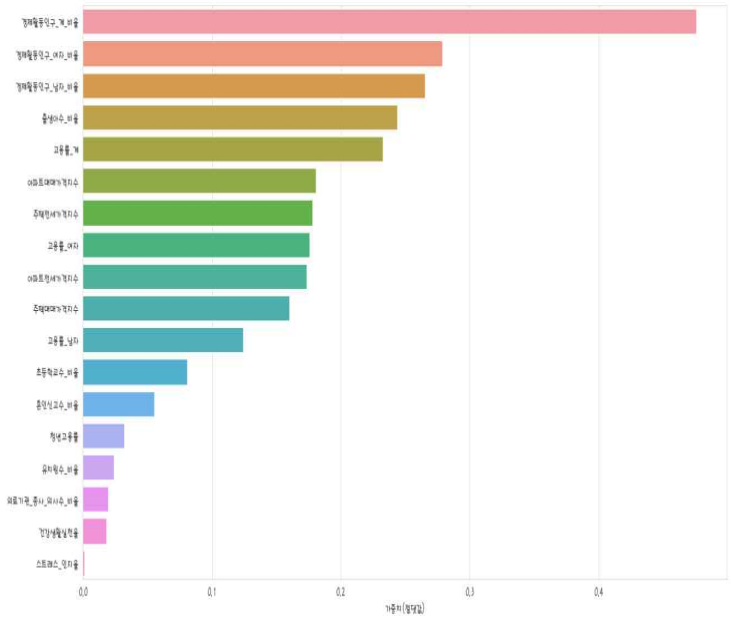

그림10. 다중회귀분석을 통한 컬럼별 가중치 값

추가적으로 Random Forest를 통해 어떤 컬럼(x)이 합계출산율(y)에 큰 영향을 끼치는지 변수 중요도를 확인해보았다. Random Forest는 의사 결정 트리를 기반으로 한 앙상블 모델이다. 의사 결정 트리는 주로 예, 아니요로 대답할 수 있는 문제에 대해 데이터를 구분해나가는 모델이다. 하지만 우리가 사용한 데이터는 수치로만 구성된 데이터이기 때문에 Random Forest 모델을 사용하기에 적합하지 않다고 판단하였다. 정리하면, Random Forest는 모델의 적합성을 설명하기 위한 용도로 변수 중요도를 사용한다. 따라서, Random Forest를 통한 변수 중요도는 통계분석에 활용하지 않았다.

마지막으로, 모델링을 통한 합계출산율(y)을 예측하였다. 모델을 설계하기 전에 다중공선성(Multicollinearity)을 먼저 확인하였다. 회귀분석에서 일부 독립변수가 다른 독립변수와 상관관계가 높아 데이터분석에 부정적인 영향을 끼칠 수 있다. 이러한 문제를 방지하기 위해 다중공선성의 파악이 필요하다. 다중공선성은 독립변수 간 상관관계가 높은지 파악하는 수치이며, 분산팽창인수(Variance Inflation Factor, VIF)로 파악할 수 있다. VIF가 10 이상일 경우에 독립변수 간 강한 상관관계를 가진다고 할 수 있으며, 회귀 모델은 신뢰가 떨어지게 된다. 데이터를 확인해 본 결과 VIF가 10 이상인 수치가 다수 확인되었다(그림11).

	VIF		VIF		VIF		VIF
Intercept	1.000000	스트레스_인지율	1.377816	주택전세가격지수	35.248169	출생아수_비율	8.797403
건강생활실천율	1.494223	청년고용률	2.890511	경제활동인구_계_비율	3826.739670	유치원수_비율	7.010905
고용률_계	1703.339502	아파트매매가격지수	24.971790	경제활동인구_남자_비율	1120.799975	의료기관_종사_의사수_비율	1.749123
고용률_남자	356.824079	아파트전세가격지수	36.681103	경제활동인구_여자_비율	1493.417126	초등학교수_비율	9.135084
고용률_여자	1023.029573	주택매매가격지수	22.916910	혼인신고수_비율	8.414969		

그림11. 모든 독립변수에 대한 VIF 수치 결과

VIF 수치가 10 이상인 독립변수들이 파악되었고, 이를 해결하기 위해 주성분 분석(Principal Component Analysis, PCA)을 진행하였다. PCA는 가장 흔하게 사용되는 차원 축소 기법의 하나다. 기존 데이터의 분포를 유지하면서 높은 차원의 데이터를 낮은 차원의 데이터로 변환해준다. PCA를 진행한 결과, 고윳값이 0.7 이상이고 누적기여율이 80% 이상 넘어가는 지점까지의 주성분인 PCA5까지 사용하였다(그림 12). PCA를 통해 차원의 문제 해결과 독립변수 간의 상관성을 해소하였고, 출산율 예측모델을 구축하였다.

	설명가능한_분산_비율(고윳값)	기여율	누적_기여율				
pca1	6.408425	0.354592	0.354592	pca10	0.218301	0.012079	0.976237
pca2	3.471256	0.192072	0.546665	pca11	0.196200	0.010856	0.987093
pca3	2.293794	0.126921	0.673586	pca12	0.087666	0.004851	0.991944
pca4	2.064511	0.114234	0.787820	pca13	0.059692	0.003303	0.995247
pca5	0.889196	0.049201	0.837021	pca14	0.050753	0.002808	0.998055
pca6	0.821472	0.045454	0.882475	pca15	0.017249	0.000954	0.999009
pca7	0.625175	0.034592	0.917067	pca16	0.012845	0.000711	0.999720
pca8	0.515016	0.028497	0.945564	pca17	0.003867	0.000214	0.999934
pca9	0.336036	0.018594	0.964158	pca18	0.001078	0.000060	0.999994
				pca19	0.000116	0.000006	1.000000

그림12. 모든 독립변수에 대한 PCA 결과

다양한 예측모델 중 XGBoost 회귀모델을 선택하였다. XGBoost는 머신러닝 예측모델에서 높은 결정계수(설명력)를 나타내는 모델 중 하나이다. 결정계수는 예측 하는 값의 정밀도 확인에 사용되는 지표로 0에서 1까지 값으로 표현된다. 수식은 다음과 같다. $R^2 = \frac{Q - Q_e}{Q}$ 여기서 Q는 전체 데이터의 편차들을 제곱하여 합한 값을 의미하며, Q_e는 전체 데이터의 잔차들을 제곱하여 합한 값이다. 결정계수가 1에 가까울수록 회귀직선과 매우 밀접하게 분포되며, 예측하는 값의 정밀도가 높음을 의미한다. 모델을 평가하기에 앞서 파라미터 값 설정이 중요하다. GridSearchCV는 머신러닝에서 모델의 하이퍼파라미터(Hyper-parameter) 값을 최적화 해주는 기법이다. 결정계수를 평가하기에 앞서 GridSearchCV를 활용해 출산율 예측 모델의 파라미터 값을 도출하였다. XGBoost를 사용한 회귀 모델이 약 75.1%의 결정계수(설명력)로 가장 높은 수치를 나타냈다(그림13). 게다가 평균 절대 오차는 약 0.09로 충분히 크지 않다고 판단되어 최종 출산율 예측모델로 선정하였고 시각화로 확인하였다.

```
xgb = XGBRegressor()
xgb_param_grid={
    'n_estimators' : [50,100,200,300],
    'learning_rate' : [0.01,0.05,0.1,0.15],
    'max_depth' : [2,3,5,7],
    'gamma' : [0,1],
    'colsample_bytree' : [0.7,0.8,0.9],
    }
# model.fit(x, y)
xgb_grid=GridSearchCV(xgb, param_grid = xgb_param_grid, scoring="r2", n_jobs=-1, verbose = 2)
xgb_grid.fit(x_train, y_train)
```

```
Fitting 5 folds for each of 384 candidates, totalling 1920 fits
```

GridSearchCV
estimator: XGBRegressor
XGBRegressor

```
print('최적의 파라미터 : ', xgb_grid.best_params_)
print('최적의 결정계수 : ', xgb_grid.best_score_)
```

```
최적의 파라미터 :  {'colsample_bytree': 0.8, 'gamma': 0, 'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}
최적의 결정계수 :  0.7512259385412279
```

```
final_xgb = XGBRegressor(colsample_bytree = 0.8, gamma = 0, learning_rate = 0.1, max_depth = 3, n_estimators = 100)
final_xgb.fit(x_train, y_train)
y_pred = final_xgb.predict(x_test)
print(f'평균절대오차 : {mean_absolute_error(y_test, y_pred)}')
```

```
평균절대오차 : 0.09149811894243415
```

```
final_xgb = XGBRegressor(colsample_bytree = 0.8, gamma = 0, learning_rate = 0.1, max_depth = 3, n_estimators = 100)
final_xgb.fit(x_train, y_train)
plot_tree(final_xgb)
plt.show()
```

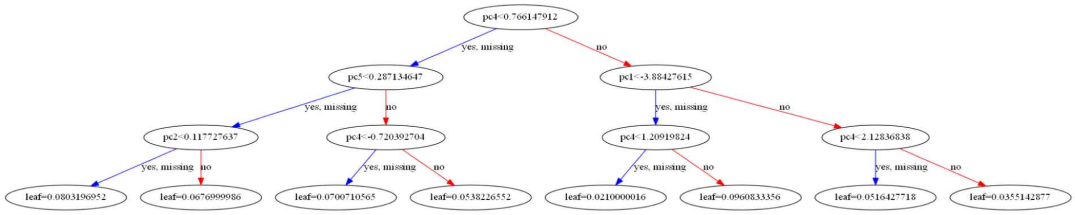

그림13. XGBoost를 활용한 모델 학습 결과 및 모델 시각화

3. 분석 활용 전략

3.1. 분석 요약

다중선형회귀를 통해 경제활동인구 비율과 주택매매가격지수 및 아파트전세가격지수가 의미 있는 가중치 값을 지닌 것을 확인하였다. 주성분분석과 XGBoost를 활용한 회귀 모델로 출산율을 예측할 수 있었다. GridSearchCV를 활용하여 하이퍼 파라미터를 최적화한 회귀 모델은 최종적으로 약 75.1%의 결정계수(설명력) 및 0.09의 평균 절대 오차 값을 가진다.

3.2. 방향성 제시

다중회귀분석을 통해 확인한 유의미한 가중치 값은 경제 활동 인구 비율(경제활동인구_계_비율)과 부동산 가격 지수(주택전세가격지수와 아파트매매가격지수)이다. 본연구의 데이터 및 분석에 의하면 경제 활동 인구 비율은 증가할수록, 부동산 가격 지수가 감소할수록 출산율이 높아지는 경향을 확인했다. 출산율을 증가시키기 위해서는 부동산 가격 지수 안정화와 경제 활동 인구 증대가 필요하다.

참고문헌

건강보험심사평가원 자원평가실, 2022, 2023.07.20, 인구 천명당 의료기관 종사 의사수(시도/시/군/구).

김우림. (2021). 저출산 대응 사업 분석·평가. 서울: 국회예산정책처.

보건복지부 인구정책총괄과, 「제4차 저출산·고령사회 기본계획」, 2020.12.31.

이삼식, 윤여원, & 이지혜. (2012). 출산에 영향을 미치는 요인에 대한 심층 사례분석.

이소영. (2023). 임신・출산 지원 정책 모니터링: 제4차 저출산고령사회기본계획을 중심으로. Health and Welfare Policy Forum, 2023(3), 7-20. https://doi.org/10.23062/2023.03.2

이재희. (2023). 임신・ 출산 및 영유아 의료 인프라 추이 분석 및 대응 방안.

양봉모, “인구학자 데이비드 콜먼 교수 "한국 저출산 지속되면 2070년 국가 소멸 위험”, BBSNews, 2023년 5월 17일자.http://news.bbsi.co.kr/news/articleView.html?idxno=3110939.

질병관리청,「지역사회건강조사」, 2022, 2023.07.20., 건강생활실천율(시도/시/군/구).

질병관리청 만성질환관리국 만성질환관리과, 2022, 2023.07.20, 스트레스 인지율(시도/시/군/구).

통계청,「2023년 3월 인구동향」, 2023.6.24.

통계청,「경제활동인구조사」, 2023.06, 2023.07.20., 경제활동인구(시도).

통계청,「경제활동인구조사」, 2023.06, 2023.07.20., 고용률(시도).

통계청,「인구동향」, 2023, 2023.07.20.

통계청,「인구동향조사」, 2022.12, 2023.07.20, 시도/시군구/월별 혼인.

통계청,「인구동향조사」, 2021, 2023.07.20., 합계출산율(시도/시/군/구).

통계청,「지역별고용조사」, 2022 2/2, 2023.07.20., 고용률(시/군/구).

통계청,「지역별고용조사」, 2022 2/2, 2023.07.20., 경제활동인구(시/군/구).

통계청,「지역별고용조사」, 2022 2/2, 2023.07.20., 청년고용률(시/군/구).

한국교육개발원, 2022, 2023.07.20., 유치원수(시도/시/군/구).

한국교육개발원, 2022, 2023.07.20., 초등학교수(시도/시/군/구).

한국부동산원, 2023.06, 2023.07.20., 주택매매가격변동률(시도/시/군/구).

한국부동산원 부동산통계처 주택통계부, 2023.05, 2023.07.20., 아파트매매가격지수(시도/시/군/구).

한국부동산원 부동산통계처 주택통계부, 2023.05, 2023.07.20., 주택매매가격지수(시도/시/군/구).

한국부동산원 부동산통계처 주택통계부, 2023.05, 2023.07.20., 주택전세가격지수(시도/시/군/구).

한국부동산원 부동산통계처 주택통계부, 2023.05, 2023.07.20., 아파트전세가격지수(시도/시/군/구).

OECD (2023), Fertility rates. last modified 2022, accesed July 10,https://data.oecd.org/pop/fertility-rates.htm.