****

**KakaoTalk_20210212_005730042**

**경사하강법**(Gradient Descent)**을 활용한**

**국민연금 고갈 시기 예측**

**팀명 : 산지니  
 팀원 : 정지환, 이다영**

목 차

[**제1장 서 론**](#_Toc33893438) **3**

[1.1. 연구배경](#_Toc33893439) 3

[1.2. 연구목적](#_Toc33893444) ... 5

1.3. 선행연구 5

[**제2장 본 론**](#_Toc33893452) **6**

2.1. 연구설계 6

2.2. 실증분석 7

2.2.1. 변수설정. 7

2.2.2. 데이터 수집 및 전처리 . 8

2.2.3. 머신러닝 모델 구축 9

2.2.4. 시나리오 예측 12

2.2.5. 개선 제도변수 적용 14

2.2.6. 결과 해석 및 비교 15

[**제3장 결 론**](#_Toc33893466) **17**

3.1. 연구요약 17

3.2. 연구의의 17

3.3. 연구 한계점 및 제언 18

**참고 문헌 19**

**부록 21**

**요 약**

|  |
| --- |
| 저출산 및 고령화와 지속적인 경제성장률의 둔화로 인해 국민연금 고갈 예측 시기가 조금씩 앞당겨 지고 있다. 뿐만 아니라 최근 코로나 19에 따른 급격한 사회변화로 기존 예측 시기와 크게 달라질 수도 있다. 이에 따라, 본 연구에서는 기존 통계적 기법을 보완하고 정확한 고갈 시기 예측을 위해 머신러닝 기법을 이용하였다. 실제 국민연금 재정계산에서 사용되는 여러 변수를 기반으로 데이터 세트를 구성하였다. 이를 경사하강법으로 반복학습 시킨 후 다중 선형회귀 모델을 구축하였다. 다양한 경우를 대비해 각기 다른 경제**·**제도적 변화를 반영하여 현상 유지, 성장 대응, 개선 3안, 개선 4안 시나리오를 가정하였다. 시나리오별로 머신러닝 모델을 통해 예측하였을 때, 보험료율과 소득대체율 인상 개선방안에 해당하는 개선 3안과 개선 4안 시나리오의 고갈 시기가 2060년, 2059년으로 가장 늦게 나타났다. 4가지 시나리오에 대한 예측 결과를 기존 연구와 비교한 결과 시나리오별 고갈 순서는 동일하며 고갈 시기는 작게는 1년, 크게는 3년 차이 밖에 나지 않았다. |

주제어: 국민연금, 고갈 시기, 머신러닝, 경사하강법, 시나리오, 예측

**제1장 서 론**

**1.1. 연구 배경**

**1.1.1. 국민연금의 적자와 고갈**

생명보험이나 손해보험과 같은 민영보험 회사들의 적립금이 고갈된다는 이야기는 없어도 사회보험 기금 고갈 여부에 대한 이슈들은 꾸준히 거론되어 왔다. 최근 8대 사회보험에 해당하는 국민**·**공무원**·**군인**·**사학(사립학교교직원) 4대 연금과 고용**·**산재**·**건강**·**노인장기요양 4대 보험의 재정전망 결과 예상보다 이른 적자전환 및 고갈 시기가 잇따라 발표되고 있다. 이러한 사회보험의 재정 문제는 보험별 운영방식 따라 정부의 사회보장지출로도 이어질 뿐만 아니라 사회보험 보장에 대한 국민들의 불안감을 증가시키고 있다.

특히, 국민연금의 재정 문제에 대한 국민들의 관심은 다른 사회보험보다 항상 클 수밖에 없다. 만19세 이상 가구주를 대상으로 한 2019년 통계청의 노후준비방법 자료에 따르면 국민연금이 58.7%를 차지한다. 국민연금 다음으로 높은 준비방법으로 예금**·**적금 저축성 보험이 15.6%, 그 외 나머지 기타 공적연금, 사적연금, 퇴직금 등은 모두 한자리 비율에 그치며 국민연금이 독보적으로 가장 높다. 이는 노후소득을 보장받기 위해 국민연금에 대한 의존성이 높다는 것을 알 수 있다.

국민연금은 기본적으로 가입자에게 유리한 수급구조로 저부담**·**고급여의 제도적인 측면을 가진다. 이러한 구조에서 초기 국민연금의 소득대체율[[1]](#footnote-1)이 높은 상태를 유지할 경우 미래 재정 지속가능성에 큰 타격을 준다. 그래서 국민연금은 국민연금법 제4조 의거하여 5년마다 향후 70년 재정전망과 그에 따른 재정개혁 방안을 마련하기 위해 재정계산제도를 시행하고 있다(국회예산정책처, 2019). 1998년 법 개정 후 2003년에 1차 재정계산이 실시되었다. 가장 최근에 실행한 4차(2018년) 재정계산 결과 2042년에 재정수지 적자가 발생하고 2057년에 기금 고갈로 이어질 것으로 전망하였다. 이는 2, 3차 재정 결과에서의 고갈 시기보다 3년 이르다.

해마다 심화되고 있는 저출산 및 고령화 현상과 불안정한 경제성장률로 인해 국민연금 고갈 예측 시기가 조금씩 앞당겨 지고 있다. 이뿐만 아니라 최근 코로나 19에 따른 급격한 사회변화로 기존 예측 시기와 달라질 수도 있다. 이에 따라 국가의 현실적인 상황을 나타내는 변수들을 통해 정확한 고갈 시기를 예측할 필요가 있다.

**1.1.2. 머신러닝 분석 기법의 활용**

최근 여러 정보통신기술(ICT) 중에서 인공지능(AI)의 발전으로 머신러닝(Machine Learning, 기계학습)이 다양한 산업 분야에 적용되고 있다. 머신러닝은 자동화와 연결성이 극대화 되는 4차 산업혁명의 핵심 기술인 인공지능의 한 분야이다. 일반적으로 컴퓨터가 대량의 과거 데이터를 기반으로 패턴을 스스로 학습하고 미래 결과를 예측(Prediction)하는 알고리즘 기법을 통칭한다.

Rajula et al.(2020)에 따르면 전통적인 통계기법은 변수 간의 관계를 추론하는 것을 목표로 하는 반면 머신러닝은 가능한 정확한 예측을 하는 데에 중점을 둔다. 복잡한 비선형 상호작용이 있는 경우에도 효과적인데(Bzdok et al., 2018) 이는 알고리즘이 현실적인 노이지(Noisy)한 데이터에 대해 가정을 하지 않기 때문이다. 이러한 이점으로 여러 실증연구에서 예측 성과가 기존 통계기법보다 높은 것으로 검증되었다.

금융**·**보험업은 타 산업보다 데이터 보유량이 많고 빠르게 증가하기 때문에 빅데이터 이용한 머신러닝을 적극적으로 활용하고 있다. 주가예측, 고객 이탈 예측, 사기탐지, 신용평가, 인수심사 등 실무적으로 활발하게 적용되고 있다.

또한, 오미애 외(2017)는 복지사각지대 예측모형과 같이 머신러닝 기법으로 사회보장 빅데이터를 분석함으로써 효과적인 정책수립 및 집행으로 공공**·**행정부문에서 효율성을 높일 수 있다고 제시하였다. 이처럼 노령, 빈곤, 사망 등 사회적 위험으로부터 국민을 보호하는 사회보험이나 공공부조와 같은 사회보장 분야에서도 머신러닝을 활용할 수 있다.

**1.2. 연구 목적**

국민연금 고갈 시기를 예측하기 위한 기존 연구들은 대부분 통계적 기법으로 진행되어 왔다. 본 연구에서는 기존 통계적 기법을 보완하기 위해 머신러닝 분석 기법을 이용한다. 이를 통해 현실적인 사회**·**경제의 비선형 데이터를 반영하면서 신뢰도 있는 고갈 시기를 예측하는 모델을 구축하고자 한다.

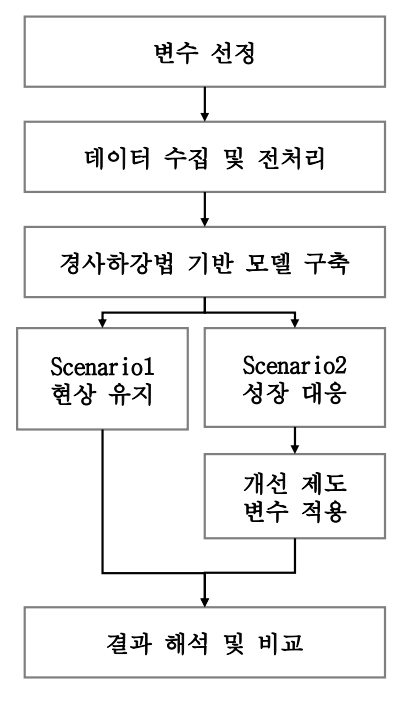
**1.3. 선행 연구**

국민연금의 장기 재정전망에 관한 기존 연구들은 많은 연구 수행기관들에 의해 진행되어왔다. 국민연금공단에서 5년 마다 실시하는 재정계산은 정부 추계이기 때문에 연구들 사이에서 기준이 되는 연구이다.

국민연금 발전위원회(2003)는 연금수리모형에 다양한 인구, 사회경제 및 정책변수를 적용함으로써 1차 재정계산 결과 2047년에 기금이 고갈될 것이라 전망하였다. 2차 재정계산에서 국민연금 재정 추계위원회(2008)는 급여율 인하와 같은 연금법 개정과 기초노령연금제도 도입 등 제도개혁으로 소진 시기를 연장할 수 있음을 확인하였다. 국민연금 재정 추계위원회(2018)는 최근 4차 재정계산으로 변수들의 조합시나리오와 개별 변수 시나리오를 통해 민감도 분석을 진행하였다. 재정 안정화를 위해서 출산율 제고와 경제발전이 중요한 요소임을 확인하였다. 또한, 다른 기관에서 연금의 재정전망에 관한 연구도 진행되어왔다. 기획재정부(2020)는 출산율 및 성장률, 제도개선 등을 고려한 4가지 시나리오에 따라 국민연금의 적자전환 시기와 고갈 시기를 예측하였다.

한편, 국내 최초로 기존 통계적 기법이 아닌 머신러닝 기법을 이용해서 국민연금 고갈 시기를 예측한 연구사례가 있다. 송영서(2020)는 심각한 저출한 현상을 고려해 연도별 출생아 수 데이터를 역전파 알고리즘에 적용하여 적립금이 0이 되는 시기를 예측하였다. 적립금이 고갈되는 이유는 출생아 수가 감소함에 따라 국민연금 가입자는 늘어나는 반면 수령자는 급격하게 감소하기 때문임을 확인하였다.

본 연구는 머신러닝 기법을 사용하되, 기존 연구와 다르게 출생아 수와 같이 단일 특정 변수가 아닌 실제 국민연금 재정계산 시 사용되는 여러 범주의 변수들을 적용하여 정확한 고갈 시기를 예측하고자 하였다. 뿐만 아니라 경제**·**제도적 변화를 반영한 시나리오를 가정하여 시나리오별 서로 다른 예측 결과를 비교하고자 하였다.

**제2장 본 론**

**<그림1> 전체 연구 과정**

**2.1. 연구 설계**

본 연구의 분석은 프로그래밍 언어로 파이썬(Python 3.7.10) 과 머신러닝에 사용되는 오픈소스 프로그램인 텐서플로우(Tensorflow 2.4.1)를 이용한다. <그림 1>은 전체 연구 과정을 나타낸다. 본 연구는 크게 6단계 별로 나누어 진행한다.

수집할 데이터의 변수를 우선적으로 선정한다. 선정한 변수들의 데이터를 수집하고 학습 데이터 세트와 테스트 데이터 세트를 각각 구성한다. 수집한 데이터 세트를 머신러닝 알고리즘에 적합하게 변환시켜주는 전처리 작업을 거친다. 그다음 앞 단계에서 구성한 학습 데이터 세트를 경사하강법으로 반복 학습시킨 후 최적의 머신러닝 회귀 예측 모델을 구축한다. 학습된 모델을 통해서 2가지 시나리오(현상 유지, 성장 대응)별로 다른 테스트 데이터를 입력 후 예측값을 도출한다. 나아가 시나리오2(성장 대응)는 개선된 제도변수 가정하에 추가적인 예측값을 도출한다. 마지막으로 전체 연구 결과를 해석하고 선행연구 결과와 비교한다.

**2.2. 실증 분석**

**2.2.1. 변수 선정**

머신러닝 기법을 사용하기 위해서는 기본적으로 입력변수와 목표변수 데이터가 필요하다. 본 연구에서의 목표변수는 연도별 국민연금 적립금 변화량으로 설정하였다. 이에 따른 입력변수는 국민연금 재정 추계모형의 가정변수를 기반으로 선정하였다. <그림2>는 국민연금연구원에서 자체개발한 장기재정추계모형의 전반적인 구조에 해당된다. 국민연금 재정계산 시 장기 전망 값으로 입력되는 가정변수(외생변수)로는 크게 인구변수, 거시경제변수, 제도변수, 기금투자수익률이 있다(국민연금 재정 추계위원회, 2018).

이를 기반으로 선정한 9개의 입력변수들 역시 4가지 범주로 나누웠다. 이는 인구입력변수, 경제입력변수, 제도입력변수, 투자수익률로 구성된다. 각 변수 범주에 대한 설명은 다음과 같다.

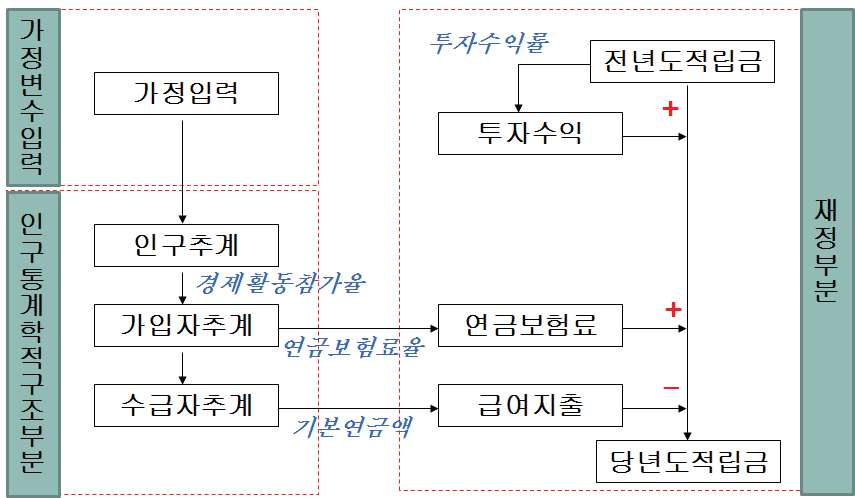
1. 인구입력변수는 출산율, 기대수명, 가입률, 이동률 등 이를 모두 반영하는 수치인 가입자와 수급자 수로 설정한다.

2. 경제입력변수에 해당하는 변수들은 재정계산에서 사용하는 거시경제 지표4가지를 이용한다. 실질경제성장률(GDP), 실질임금상승률, 실질금리, 물가상승률이 이에 해당한다.

3. 제도입력변수는 재정계산 시 연금보험료 수입과 급여지출에 적용되는 보험료율과 소득대체율로 설정한다.

**출처: 국민연금 재정추계 위원회, 국민연금**

**장기재정추계, 2018**



**<그림2> 국민연금 장기재정정추계모형 구조**

4. 투자수익률은 기금투자수익률에 해당한다. 이는 국내외 주식 및 해외투자의 확대 등 투자 다변화에 따른 성과제고를 나타낸다.

**2.2.2. 데이터 수집 및 전처리**

**1) 데이터 수집**

머신러닝 알고리즘은 과거 데이터를 학습하고 이를 기반으로 미래를 예측한다. 따라서 모델을 학습하기 위한 학습 데이터와 예측을 위한 테스트 데이터 모두 필요하다. 각 데이터 세트에 대한 내용은 다음과 같다.

**<표1> 학습 데이터 변수표**

1. 학습 데이터 세트

본 연구는 선정한 입력변수와 목표변수에 대해 1999년부터 2019년까지 21년간의 데이터를 수집하고 이를 학습데이터 세트로 직접 구성하였다. 이때 목표변수는 적립금 변화량이다. <표 1>은 전체 학습 데이터의 변수와 출처를 나타내는 표이다.

|  |  |  |
| --- | --- | --- |
| **변수명** | **변수 설명** | **출처** |
| Subscribe | 가입자 수 | 국민연금공단  연금 통계 |
| Beneficiary | 수급자 수 |
| Real GDP\_  GrotwthRate | 실질 경제성장률 | 한국은행 국민소득통계 |
| Wage\_GrotwthRate | 실질 임금상승률 | 고용노동부 임금결정현황조사 |
| Interest\_Rate | 실질 금리[[2]](#footnote-2) | 한국은행 경제통계시스템 |
| Inflation\_Rate | 물가 상승률 | 통계청 소비자물가지수 |
| ReturnOnFund Investment | 기금투자수익률 | 국민연금공단연금통계 |
| Premium\_Rate | 보험료율 |
| IncomeSub\_  Rate | 소득대체율 |
| Difference Of Reserve | 적립금 변화량 |

2. 테스트 데이터 세트

본 연구에서 테스트 데이터 세트는 10년 단위로 2020년부터 2060년까지의 입력변수 데이터로만 구성되어있다. 이는 2020년~2060년에 해당하는 국민연금 적립금 변화량(목표변수)은 학습된 머신러닝 모델이 예측하는 값이기 때문이다. <표 2>는 전체 테스트 데이터의 변수와 출처를 나타내는 표이다.

|  |  |  |
| --- | --- | --- |
| **변수명** | **변수 설명** | **출처** |
| Subscribe | 가입자 수 | 기획 재정부,  2020~2060년  장기재정전망,  2020 |
| Beneficiary | 수급자 수 |
| Real GDP\_  GrotwthRate | 실질 경제성장률 |
| Wage\_GrotwthRate | 실질 임금상승률 | 국민연금  재정추계  위원회,  국민연금 장기재정추계,  2018 |
| Interest\_Rate | 실질 금리 |
| Inflation\_Rate | 물가 상승률 |
| ReturnOnFund Investment | 기금투자수익률 |
| Premium\_Rate | 보험료율 | 국민연금공단연금통계 |
| IncomeSub\_  Rate | 소득대체율 |

**<표2> 테스트 데이터 변수표**

<표1>의 학습 데이터와 <표2>의 테스트 데이터 출처를 보면 큰 차이가 존재한다. 학습 데이터는 과거의 실제 통계치고 테스트 데이터의 경우 미래 전망 값에 해당한다.

신뢰성 확보를 위해 테스트 데이터는 2018년 에 시행된 4차 재정계산과 2020년에 발표된 기획재정부의 장기재정전망에서 활용한 전망 값들로 구성하였다. 기획재정부의『 2020~2060년 장기 재정전망[[3]](#footnote-3)』은 정부 차원에서 실시한 가장 최근의 재정전망이다. 이에 최근 코로나 19위기를 반영한 경제성장률과 저출산**·**고령화 현상에 따른 가입자, 수급자 수를 전망하였다. 본 연구는 이를 테스트 데이터로 활용하였다.

보험료율과 소득대체율을 제외한 나머지 입력변수 테스트 데이터는 2018년 4차 재정계산 시 전망했던 값을 이용하였다. 제도입력변수인 보험료율과 소득대체율은 제도개혁 시 모두 정해지기 때문에 국민연금 통계치를 참고하였다.

**2) 데이터 전처리**

서로 다른 입력변수 데이터들의 크기를 통일하기 위해 Z-Score표준화(Standardization) 방법으로 데이터 스케일링(Scaling)을 하였다. 표준화는 Input을 가중시키는 경사하강법과 같이 최적화 알고리즘에 유용하다(Liu, 2020). 입력변수의 데이터를 평균이 0이고 표준편차가 1인 정규 분포가 되도록 변환해준다.

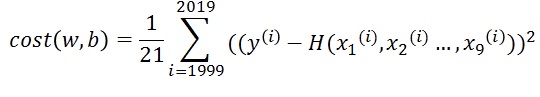
**2.2.3. 머신러닝 모델 구축**

본 연구의 데이터 세트에서 목표변수는 연속형 숫자 값인 적립금 변화량이고 예측값 역시 미래 특정 연도의 적립금 변화량이다. 이는 머신러닝 지도학습 중에서도 회귀(Regression)에 해당된다.

머신러닝 회귀예측의 핵심은 학습을 통해서 최적의 회귀계수와 절편을 찾고 회귀선을 도출하는 것이다. 앞에서 선정한 입력변수는 9개이므로 이는 다중 회귀에 해당된다. 다중 회귀 식은 다음과 같다.

(1)

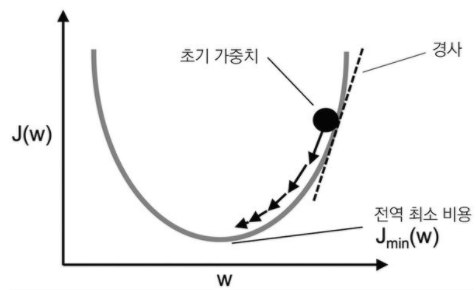
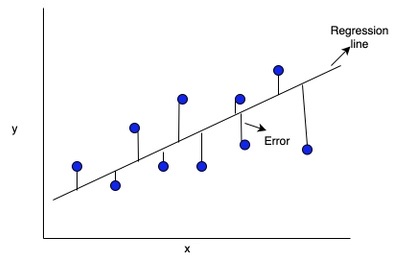
이때 식(1)에서 학습으로 도출되는 회귀계수 W1, W2… W9가 선형결합으로 표현되기 때문에 본 연구의 머신러닝 회귀 모델은 다중 선형회귀 모델이다. <그림 3>은 기본적인 단순선형회귀 그래프이다.

<그림 3>에서 학습 데이터를 잘 반영하는 최적 회귀선을 구하려면 실제값(y)과 회귀 예측값의 차이인 잔차(error)값의 합이 최소가 되어야한다. 이 때, 부호가 다른 오류 값들이 서로 상쇄되지 않고 계산을 편리하기 위해 제곱 합으로 구한다. 전체 오류 합의 제곱을 총 학습 데이터 건수로 나누면 비용함수(cost)가 된다. 식(2)는 1999년부터 2019년까지의 학습 데이터에 대한 비용함수의 수식이다.

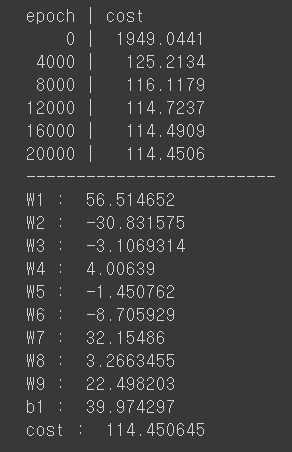
**<그림4> 경사하강법 알고리즘**

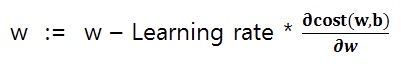
(2)

식(2)의 비용함수에서 변수는 입력되는 학습데이터가 아니라 W1, W2…W9와 b로 정의된다. 따라서 비용함수가 최소가 되는 w와 b(매개변수)값을 찾아야 데이터에 가장 적합한 회귀선을 도출할 수 있다(Pramoditha, 2020)

본 연구에서 비용함수를 최소화하는 w와 b 값을 찾기 위해 경사하강법(Gradient Descent)을 활용하였다.  경사하강법은 가장 일반적인 최적화 방법이며 변수가 반복해서 목표 함수 기울기의 반대 방향으로 업데이트되는 아이디어를 가진다(Sun, 2019). <그림 4>는 경사하강법의 작동방식을 전반적으로 나타내는 그래프이다.  접선의 기울기(경사)가 최소인 점이 비용함수가 최소인 지점이다. 따라서 그 지점의 W와 b값을 찾아야 한다. 경사하강법의 단계적 절차는 다음과 같다.

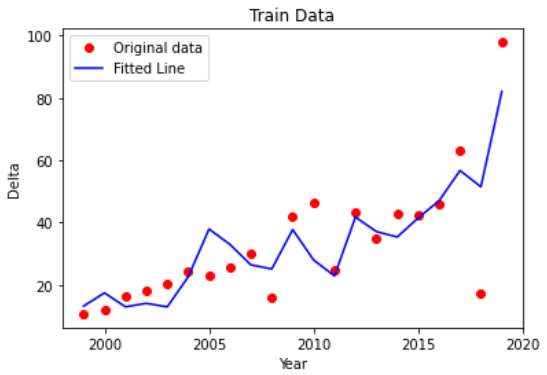
**<그림3> 단순선형회귀**

첫째, W1, W2…W9와b 에 대한 시작 값(시작점)을 임의로 설정하고 첫 비용함수의 값을 계산한다.

둘째, 비용함수를 w에 대해 편 미분한 값을 빼주는 방법으로W1, W2…W9를 업데이트 한다. b 값 역시 같은 방법으로 업데이트 한다. w를 업데이트 하는 식은 다음과 같다.

식(3)

이때 Learning rate는 편미분 값을 보정해주는 학습률이다. 업데이트한 새로운 w와 b 값을 이용해서 다시 비용함수를 계산한다.

셋째, 앞의 두 번째 단계를 입력 자로 받은 횟수만큼 반복한다.

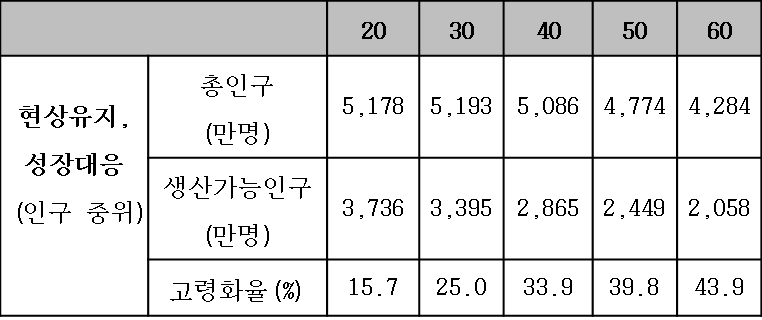
본 연구에서는 Learning rate를 0.01로 설정하고 20000번 반복 학습하여 W1, W2…W9와b 값 및 비용함수의 값을 도출하였다. <그림 5>는 실시한 경사하강법의 결과를 나타낸다. w는 회귀계수이면서 입력변수에 대한 가중치이다. 도출한 w 값을 통해서 각 입력변수의 영향력을 파악할 수 있다. <그림 6>에서 빨간 점은 학습 데이터의 국민연금 적립금 변화량이고 파란 실선이 학습을 통해 도출한 모델의 다중 선형 회귀선에 해당된다.

**<그림6> 다중 선형회귀 모델**

**<그림5> 경사하강법 결과**

**2.2.4. 시나리오 예측**

**1) 시나리오 가정**

본 연구는 경사하강법으로 구축한다중 선형회귀 모델을 이용해 서로 다른 2가지 시나리오에 대한 예측값(국민연금 적립금 변화량)을 도출하였다. 시나리오에 대한 기본 가정은 기획재정부의 『2020~2060년 장기재정전망』에서 사용한 시나리오를 기반으로 하였다. 각 시나리오의 가정은 다음과 같다.

**출처: 기획재정부,2020~2060년장기재정전망,2020**

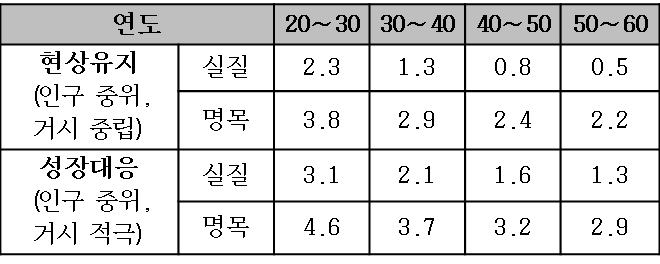
**출처: 기획재정부,2020~2060년장기재정전망,2020**

**<표 4> 시나리오별 인구 전망**

1. 시나리오 1 : 현상 유지

**<표 3> 시나리오별 성장률 전망**

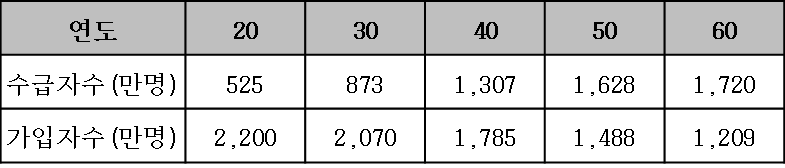
**단위 : %**

인구 중위[[4]](#footnote-4)와 거시 중립[[5]](#footnote-5)을 주요전제로 한다. 정책 무대응에 따라 저출산 등으로 장기간 인구감소의 추세와 경제성장률 둔화가 지속되는 경우이다.

2. 시나리오 2 : 성장 대응

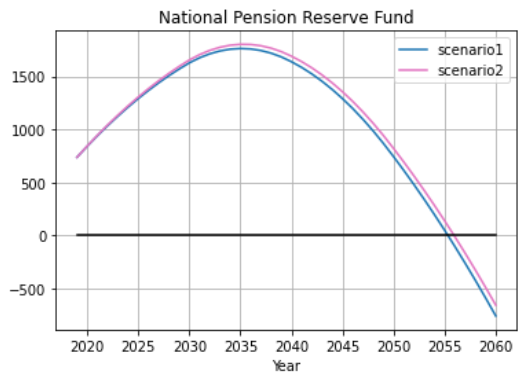
인구 중위와 거시 적극을 주요전제로한다. 인구 가정은 시나리오1과 동일하다. 4 차 산업혁명 등에 따라 총 요소 생산성이 향상되어 성장률 하락세가 완화된 경우이다.

<표 3>과 <표 4>는 『2020~2060년 장기재정전망』에서 통계청과 KDI지표를 활용한 시나리오별 성장률과 인구 전망값이다.

 『2020~2060년 장기재정전망』에서 국민연금 재정 추계를 위해 <표4> 의 인구 중위에 따른 국민연금 가입자와 수급자 수를 전망하였다. 2020~2060년 가입자와 수급자 수를 전망한 값은 다음과 같다.

**<표 5> 국민연금 수급자·가입자수 전망**

**출처: 기획재정부,2020~2060년장기재정전망,2020**

고령화율이 커지면서 수급자 수는 증가하지만 반대로 생산가능인구의 감소에 따라 가입자 수는 감소하리라 전망하였다. 이에 따라, 본 연구는 테스트 데이터로 <표 3>과 <표 5>의 전망 값을 활용하였다.

**<그림8> 국민연금 예상 적립금 (시나리오1,2)**

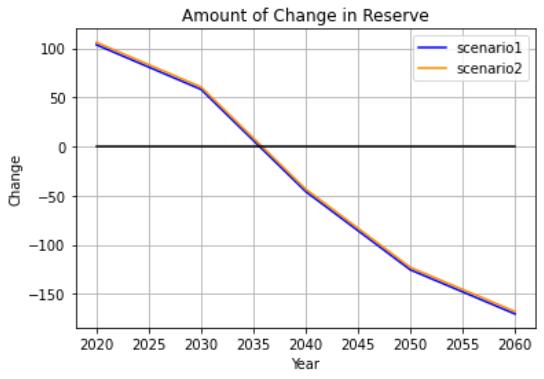
**2) 시나리오 예측**

현상 유지(시나리오1)와 성장 대응(시나리오2)에 해당하는 각각의 테스트 데이터를 구축한 다중선형회귀 머신러닝 모델로 각각 예측하였다. <그림 7>는 예측한 국민연금 적립금 변화량을 그래프로 나타낸 것이다.

변화량이 0보다 크다면 적립금이 증가함을 의미하고 0보다 작은 경우 적립금이 감소함을 의미한다. <그림 7>에서 현상 유지와 성장 대응 선은 거의 겹치면서 적자전환 시기도 모두 2035년으로 나타난다.

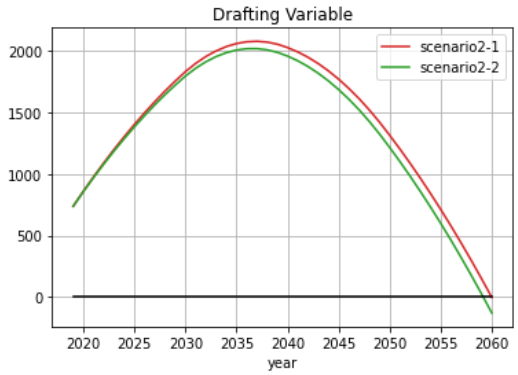
**<그림7> 국민연금 예상 적립금 변화량 (시나리오1, 2)**

<그림 8>은 예측 적립금 변화량으로 계산한 국민연금 예상 적립금이다. 보건복지부에 따르면 2019년 말 국민연금 기금적립금은 현재 시가 기준 737조 원이다. 이를 시작 적립금으로 계산하였다. 현상 유지는 2035년에 적립금이 1762조 원까지 증가하고 2055년에 고갈될 것이라 나타난다. 성장 대응의 경우 2035년에 1802조 원까지 증가하고 현상 유지보다 1년 연장된 2056년에 고갈이 예상된다

.

**2.2.5. 개선 제도변수 적용**

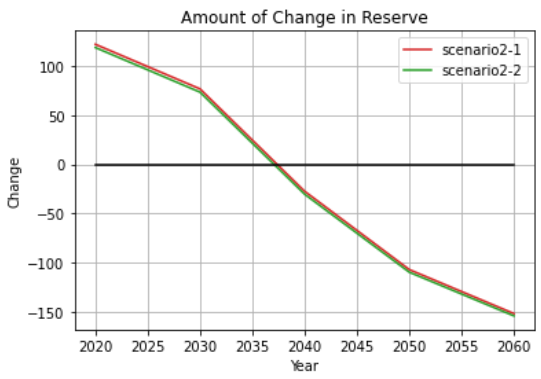
보험료율 또는 소득대체율의 인상여부는 국민연금 제도개혁의 핵심 내용이다. 보건복지부(2018)는 4차 재정계산 결과와 개선방안을 토대로 국민 의견을 수렴 후 4차 국민연금 종합운영계획을 발표하였다. 기획재정부(2020)는 종합운영계획의 정책 개선방안 3,4안을 성장 대응 (시나리오2)에 적용하여 연금 추계를 하였다.

이에 본 연구에서는 성장 대응에 개선방안 3안을 반영한 시나리오 2-1와 4안을 반영한 시나리오 2-2로 나누어 예측하였다. 각 시나리오 내 개선 내용은 다음과 같다.

**<표 6> 정책 개선방안**

**<그림9> 국민연금 예상 적립금 변화량 (시나리오2-1,2-2)**

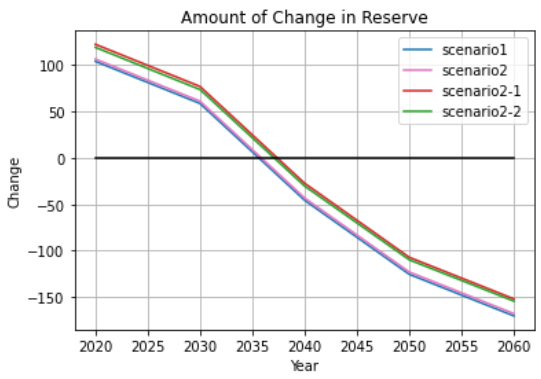
|  |  |  |  |
| --- | --- | --- | --- |
| **제도**  **변수** | **현행** | **시나리오**  **2-1** | **시나리오**  **2-2** |
| 보험료율 | 9% | 31년 12%  (21년부터 5년마다  1% 씩 인상) | 36년 13%  (21년부터 5년마다  1% 씩 인상) |
| 소득  대체율 | 28년까지  40%로 인하 | 21년부터 45% | 21년부터50% |



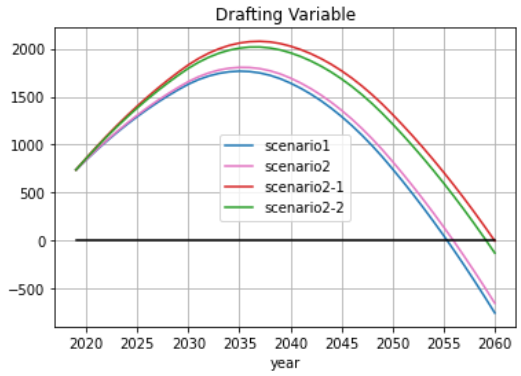
**<그림10> 국민연금 예상 적립금 (시나리오2-1,2-2 2)**

<그림 9>과 <그림 10>는 개선 3안(시나리오2-1)과 개선 4안(시나리오2-2)에 대한 예측 결과를 나타낸다. 두 시나리오 모두 적자전환 시기가 2038년으로 나타난다. 개선 4안은 적자전환 전 2037년에 2017조 원으로 적립금 최고액을 기록한다. 고갈은 2059년으로 예상된다. 개선 3안은 2037년에 2076조 원까지 증가하며 개선 4안보다 1년 연장된 2060년에 고갈이 예상된다.

**2.2.6. 결과 해석 및 비교**

<그림 10>과 <그림 11>은 앞서 시나리오별로 예측한 결과들을 한꺼번에 보여주는 그래프이다.

**<그림 11> 전체 시나리오 국민연금 예상 변화량**

<표 5>에서 보여주듯이 가입자 수는 감소하고 수급자 수가 증가하는 추세가 지속되면 저부담**·**고급여의 수급구조에 따라 기금 적립금은 더 빨리 고갈될 것이다. 이에 따라, 인구 중위 기준으로 재정 안정화를 유지하기 위해서는 현재 경제성장률 하락폭이 기본적으로 완화되어야 한다.

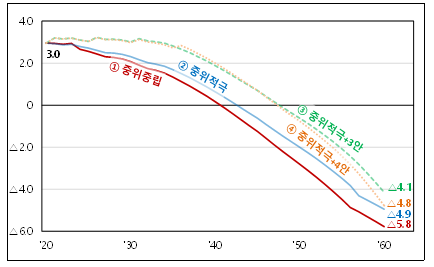
**<그림 10> 전체 시나리오 국민연금 예상 변화량**

한편, <그림 11>에서 개선 제도변수를 적용하기 전과 후의 적립금 차이가 확연하게 난다. 본 연구의 예측 결과로는 개선3안이 시나리오 중에서 고갈 시기가 2060년으로 가장 늦다. 국민연금의 재정 안정성을 제고하기 위한 개선방안을 모색하는 것은 중요한 부분이다. 하지만, 국민 연금은 국민의 노후소득보장을 목표로 도입된 사회보장제도이기 때문에 재정 안정화와 노후소득보장의 조화로운 설계가 요구된다(최장훈, 2018). 보건복지부(2018)에 따르면 3, 4안 모두 노후소득보장 강화방안에 해당된다.

따라서 개선3안과 같이 소득대체율을 높게 인상함으로써 노후소득보장을 강화하는 동시에 기금의 재정안정화를 위해 장기간에 걸친 단계적인 보험료율 인상을 추진하는 개선방안을 도입할 필요가 있다.

**<표 7> 기존연구와 결과 비교**

**단위 : 년**

****<표 7>는 기획재정부의 국민연금 재정 추계 결과와 본 연구 결과를 비교한 표이다. 본 연구에서 머신러닝 회귀 모델이 예측한 미래값의 정답을 아직 모르기 때문에 정확한 성능 평가가 어렵다. 따라서 기존 연구와 분석 기법 및 사용한 변수에서 차이가 조금 있지만, 이 점을 감안하여 예측 결과를 서로 비교하였다.

**출처: 기획재정부,2020~2060년장기재정전망,2020**

**<그림 12> 기존연구의 시나리오별 국민연금 예상 변화량**

**<그림 12> 기존연구의 시나리오별 국민연금 예상 변화량**

고갈 시기는 가장 큰 차이가 3년이지만 적자 시기에는 큰 차이를 보인다. 적자 시기의 차이 발생 이유는 명확히 알 수 없다. 두 연구의 결과가 거의 유사한 고갈 시기는 개선 3안, 4안에 있어서만 3년 차이가 난다. 이는 <그림 5>에서 보험료율의 가중치(w8)가 대략 3으로 모델 학습 시 영향을 거의 주지 않아서 현상 유지와 성장 대응보다 큰 차이를 보이는 것이라 추정된다.

결론적으로 기존 연구와 본 연구에서 시나리오별 고갈 순서는 모두 동일하며 고갈 시기에서 작은 차이를 보였다.

**제3장 결 론**

**3.1. 연구 요약**

본 연구는 기존 통계적 기법을 보완하기 위해 머신러닝 분석 기법을 이용해 국민연금 고갈 시기를 예측하였다. 실제 국민연금 재정계산 시 사용하는 여러 변수를 기반으로 경사하강법을 이용하여 머신러닝 회귀 모델을 구현하였다. 현상 유지, 성장 대응, 개선 3안, 개선 4안 시나리오를 가정하고 예측하였다. 본 연구에서는 소득대체율과 보험료율을 개선하는 개선4안 시나리오가 고갈 시기가 2060년으로 가장 늦게 나타났다.

모델의 정확한 성능을 평가할 수 없어서 본 연구의 기반이 되는 기획재정부의 고갈 시기 예측값과 비교를 하였다. 그 결과 시나리오별 고갈 순서는 모두 동일하며 고갈 시기에서 작은 차이를 보였다. 기획재정부에서 예측한 고갈 시기는 현상 유지, 성장 대응의 경우 1년 정도 차이로 예측 결과가 유사하다. 개선3안, 4안은 동일하게 3년 차이가 났다. 학습데이터에서 보험료율은 9%로 일정하다. 이에 따라 가중치(W8)가 대략 3으로 작게 학습되어 개선된 보험료율을 적용하더라도 충분히 반영되지 못한 것으로 예상된다.

**3.2. 연구 의의**

본 연구는 인공지능 기술인 머신러닝 알고리즘을 활용해 국민연금의 기금 적자 및 고갈 시기를 서로 다른 시나리오별로 예측하였다는 점에서 큰 의의가 있다.

기본적으로 머신러닝 모델을 통해 복잡한 재정 추계 과정을 거치지 않고 기존의 통계적 기법에 의한 연구들과 유사한 고갈 시기를 도출할 수 있다. 또한, 각기 다르게 가정한 시나리오를 머신러닝 모델로 예측**·**비교함으로써 국민연금의 재정 안정화와 국민들의 노후소득보장을 위한 조화로운 개선방법에 대해 고찰할 수 있다.

**3.3. 연구 한계점 및 제언**

본 연구가 가지는 한계점은 다음과 같다.

1. 학습 데이터의 부족

기본적으로 선정한 변수들에 대한 데이터 수집에 어려움이 있었다. 이에 본 연구는 1999년부터 2019년까지의 21년간의 데이터를 수집하여 모델을 학습시키고 향후 40년을 예측하였다. 머신러닝 모델을 학습시키기에는 데이터 수가 매우 작기 때문에 모델 학습시키는 데에 한계가 있었다.

2. 예측 데이터의 한계

본 연구에서 예측하고자 하는 출력 변수는 적립금 변화량이다. 이를 위해 사용되는 입력 변수로 보험료율, 소득대체율을 제외한 모든 변수들은 미래의 출생아 수, 생산 인구수, 고령화율을 전망한 자료를 토대로 예측한 값들이다. 입력 변수가 예측을 통해 설정된 값이기 때문에 입력 변수의 예측 정확도에 따라 출력 변수의 적립금 변화량 정확도가 좌우된다.

3. 제한적인 변수

실제 국민연금 재정 추계를 할 때는 인구, 경제, 제도 부분에서의 다양한 변수들을 고려하여 국민연금 재정관련 사항들을 예측한다. 배경지식의 부족으로 복잡한 추계방법을 완전히 이해하는데는 어려움이 있었기 때문에 본 연구에서는 각 부분을 대표하는 대표 변수들만 추출하였다. 그에 따른 한계로 적립금 고갈과 같은 전체적인 흐름만 파악할 수 있고 모든 부분에서 세부적으로 정확한 예측을 하기 어려웠다. 따라서, 향후 연구에서 재정계산 변수를 최대한 많이 반영하여 연구를 진행한다면 더 정확한 예상치로 다른 연구기관들의 결과와 동등한 비교할 수 있을 것이다.

4. 성장 대응만 가정

본 연구에서는 인구 고위의 경우를 배제하고 인구 중위를 기준으로 고령화가 지속됨을 가정하고 진행하였다. 하지만 출산율 제고, 외국인 인력 활용 등에 따라 인구 감소 폭이 둔화되고 고령화율이 감소하는 상황을 완전히 배제할 수는 없다. 향후 연구자들은 인구 고위의 시나리오를 가정을 추가하여 국민연금 재정을 추계한다면 다양한 관점에서의 전망을 얻을 수 있을 것이다.

**참고 문헌**

[1] 국회예산정책처.2019.2019~2060년 국민연금 재정전망.서울:국회예산정책처.

[2] 국회예산정책처.2020.4대 공적연금 장기 재정전망.서울:국회예산정책처.

[3] 기획재정부.2020.2020-2060장기재정전망.서울:기획재정부

[4] 조현숙. 2020. ”국민연금 2054년 고갈, 8대사회 보험 다 빨간불.”중앙일보.2020.06.22.

<https://news.joins.com/article/23806910>

[5] 보건복지부.2018.제4차 국민연금 재정계산을 바탕으로 한 국민연금 종합운영계획. 서울:보건복지부.

[6] 보건복지부.국민연금재정추계위원회.2003.1차 국민연금 재정계산 보고서. 서울:보건복지부.

[7] 보건복지부.국민연금재정추계위원회.2008.2차 국민연금 재정계산 보고서. 서울:보건복지부.

[8] 보건복지부.국민연금재정추계위원회.2013.3차 국민연금 재정계산 보고서. 서울:보건복지부.

[9] 보건복지부.국민연금재정추계위원회.2018.4차 국민연금 재정계산 보고서. 서울:보건복지부.

[10] 김춘순. (2010). 4 대 공적연금 재정수지의 전망과 대책. 보건복지포럼, 2010(10), 79-88.

[11] 권혁진, & 류재린. (2018). 노후소득보장을 위한 공적연금의 적정성과 재정안정성에 대한 전망. 공공사회연구, 8(1), 38-82.

[12] 권철민.2020.파이썬 머신러닝 완벽가이드.위키북스.

[13] Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. Medicina, 56(9), 455.

[14] Bzdok, D., Altman, N., & Krzywinski, M. (2018). Points of significance: statistics versus machine learning.

[15] 최창희, & 홍민지. (2018). 빅데이터 활용 현황과 개선 방안. 보험연구원 연구보고서, 2018(29), 1-128.

[16] 오미애, 최현수, 김수현, 장준혁, 진재현, & 천미경. (2017). 기계학습 (Machine Learning) 기반 사회보장 빅데이터 분석 및 예측모형 연구. 한국보건사회연구원,

[17] 최장훈, & 강성호. 2018. 국민연금지속가능성과 제도 발전 방향. 보험연구원, 1-9.

[18] 원종현. (2010). 국민연금기금 제도 지속성을 위한 기금운용 방향. 비판과 대안을 위한 사회복지학회 학술대회 발표논문집, 276-304.

[19] 박상현. (2019). 국민연금기금의 재원 확충 방안 검토: 사행산업 수익의 활용을 중심으로. 연금연구, 9(2), 29-50.

[20] 윤석명. (2011). 인구고령화를 반영한 공적연금 재정전망과 정책과제: 국민연금과 기초노령연금 중심으로. 보건복지포럼, 2011(8), 16-27.

[21] 백혜연. (2016). 국민연금 장기재정추계 결과의 신뢰성 제고 방안: 미국 OASDI 와 캐나다 CPP 의 확률론적 재정추계모형소개중심으로. 보건복지포럼, 2016(4), 21-33.

[22] 백혜연. (2014). 국민연금 보험료율 인상 시나리오별 국민연금 장기 재정 상태 분석. 한국재정학회 학술대회 논문집, 1-14.

[23] 송영서. (2020). 역전파 알고리즘을 이용한 대한민국 국민연금 고갈 시기 예측 연구. 전자공학회논문지, 57(7), 37-43.

[24] Clare Liu.2020.” Data Transformation: Standardization vs Nomralization”.KDnuggests.

<https://www.kdnuggets.com/2020/04/data-transformation-standardization-normalization.html>

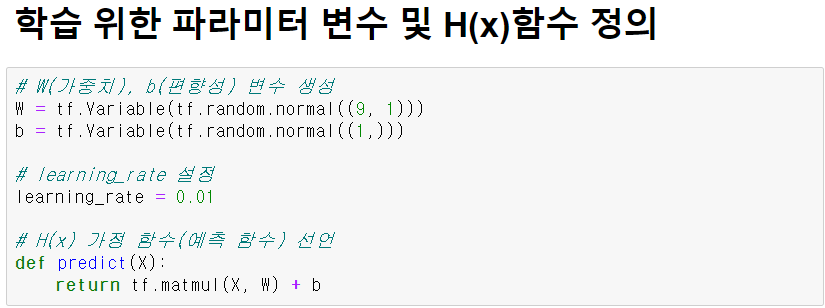
[25] Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics, 50(8), 3668-3681.

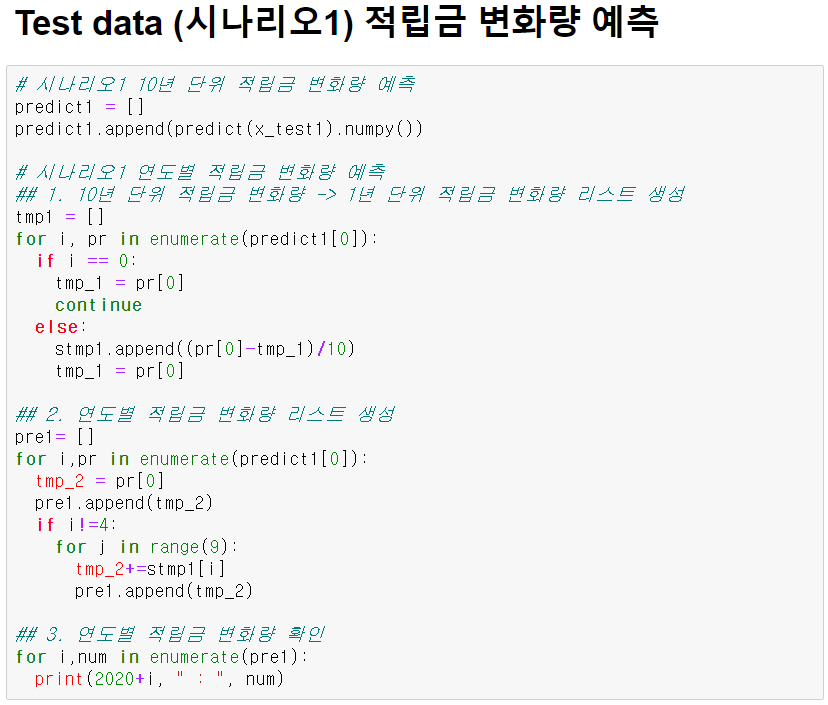
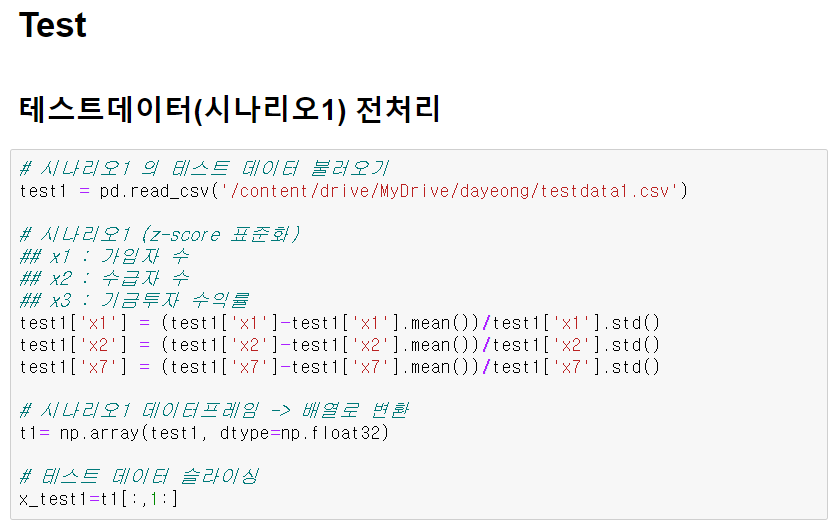
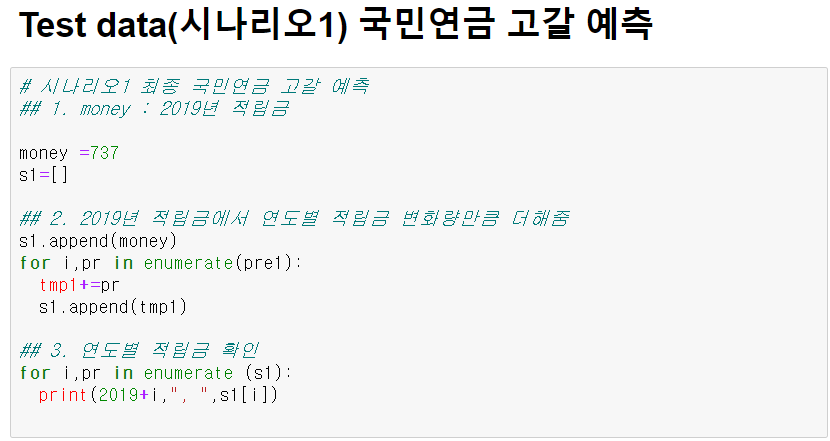
**부 록**

<부록 1> 학습 데이터 입력변수 시각화

|  |  |
| --- | --- |
| **국민연금 가입자 수** | **국민연금 수급자 수** |
|  |  |
| **실질 경제 성장률** | **실질 임금 상승률** |
|  |  |
| **실질 금리** | **물가 상승률** |
|  |  |
| **투자 수익률** | **적립금 변화량** |
|  |  |

<부록 2> Python(파이썬) 분석 코드





1. 연금급여를 가입 기간 생애평균소득의 현재 가치로 나눈 값. [↑](#footnote-ref-1)
2. 실질금리 = 회사채(3년, AA-) 금리 - 물가상승률. [↑](#footnote-ref-2)
3. 재정전망 시 통계청 장래인구특별추계(‘19.3월) 및

   KDI 거시전망지표 (’20.4월) 를 전망값으로 활용하였다. [↑](#footnote-ref-3)
4. 중간 수준의 출산율과 기대수명 및 국제순이동 조합. [↑](#footnote-ref-4)
5. 거시변수의 중립 전망치. [↑](#footnote-ref-5)