

제1회 경남소프트웨어경진대회

[딥러닝을 활용한 주택가격지수 예측]개발 완료 보고서

소속 : [Drink Coffee Like Water]

버전번호 : [Version 1.0]

제출일자 : [2020-10-15]

목 차

1. 개발 개요.....	3
1.1 개발 명	3
1.2 개발 기간.....	3
1.3 개발 목적.....	3
1.4 개발 범위.....	3
1.5 개발 기대효과.....	4
2. 개발환경	4
2.1 개발환경	4
2.2 운영환경	5
3. 개발일정	5
4. 문제분석	5
4.1 목표	6
5. 시스템 명세	6
5.1 구성	6
5.2 USE-CASE 다이어그램.....	7
5.3 CLASS 다이어그램.....	8
5.4 API 서버.....	9
5.5 GUI	9
6 학습모델 명세	11
6.1 데이터 선정.....	11
6.2 데이터 수집.....	12
6.3 학습모델 선정.....	13
6.4 데이터 전처리.....	13
6.5 전이학습과 미세조정을 사용한 모델 성능 최적화	14
6.6 예측성능	15
7 참고문헌.....	18
8 부록.....	19

1. 개발 개요

1.1 개발 명

제1회 경남소프트웨어경진대회 [딥러닝을 활용한 주택가격지수 예측] (이하 “프로그램 개발”이라 칭한다.)

1.2 개발 기간

2021년 09월 16일 ~ 2021년 10월 15일

1.3 개발 목적

- 1) 머신러닝 관련 지식 없이도 딥러닝 모델을 활용한 미래 주택가격지수를 정량적으로 예측하고 시각화 할 수 있는 프로그램을 제작하여 창원시 주택정책 수립 시 도움이 될 수 있는 참고자료로 활용
- 2) 주택가격지수의 등락과 연관성이 있다고 검증된 여러가지 경제지표 데이터를 학습하여 다방면 경제상황을 반영하고 신뢰성 있는 주택가격지수를 예측하는 학습모델 설계

1.4 개발 범위

- 1) 개발 명: [딥러닝을 활용한 창원시 주택가격지수 예측]
 - 시계열 경제지표 데이터를 학습하여 미래 주택가격지수를 예측하는 딥러닝 모델 제작
 - 학습된 모델을 의존성이 없고 누구나 간편하게 사용할 수 있도록 GUI 기반프로그램 제작
- 2) 개발내용 1
 - 창원시 주택가격지수 데이터 수집
 - 주택가격지수와 연관성이 있는 경제지표 데이터 선정 및 검증
- 3) 개발내용 2
 - 예측모델의 실효성 문의(환경도시국 주택정책과 유선문의)
 - 공공기관 협조(창원시, 한국은행, 공공데이터 포털) 및 데이터 수집
- 4) 개발내용 3

소프트웨어개발완료보고서

- 시계열 데이터 학습모델 선정 및 설계
- 학습모델 테스트
- 학습성능 개선을 위한 최적화 기법 적용(전이학습)

5) 개발내용 4

- 학습모델 직렬화
- 학습된 모델을 사용하여 데이터를 예측하기 위한 인터페이스 구현(GUI)
- 프로그램 경량화, 속도개선을 위해 GUI프로그램과 학습모델 분리
- GUI(로컬)↔학습모델(웹서버) 통신을 위한 POST API 구축
- 의존성 없이 동작할 수 있도록 프로그램 패키징

1.5 개발 기대효과

- 1) 주택정책 수립 시 참고 데이터로 활용하여 정책 결정에 도움
- 2) 다양한 경제지표 데이터를 학습한 예측데이터를 제공하여 여러가지 경제상황을 반영한 주택 정책수립 가능
- 3) 수학적으로 계산한 객관적인 예측데이터를 신속하게 제공하여 빠르게 변화는 시장상황 대응 및 통제에 용이

2. 개발환경

2.1 개발환경

구분		Spec.
시스템 환경 (자체 개발 서버)	OS	Ubuntu 16.04LTS, Window 10 Pro
	CPU	Intel Core-i5 9500 3.2GHz
	Memory	DDR4 32GB
	Storage	NVMe m.2 512GB SSD,
	개발방법	개발언어 Python3.8
	Data Base	-

소프트웨어개발완료보고서

2.2 운영환경

구분	Spec.
시스템 환경	OS Window 10 이상
	CPU Intel Core-i5 9500 3.2GHz 이상
	Memory DDR4 4GB 이상
	Storage SATA, MVM SSD 128GB 이상
	개발방법 Python3.8
	Data Base -

3. 개발일정

	1W	2W	3W	4W	공수(MM)	비고
착수	 착수				0.1	-
분석/설계		 학습데이터 수집 및 분석	 학습모델 설계		0.3	-
개발			 GUI구현 모델구현 웹서버 API구현		0.4	-
단위/통합 테스트					0.1	-
Go live 및 안정화					0.1	-
참여인원 (명)	1	1	1	1		

4. 문제분석

최근 5년동안 전국 주택가격은 꾸준히 상승세를 보이고 있다. 주택가격의 급등은 주택 실수요자, 특히 신혼부부 같은 젊은 세대의 주거불안정과 소비심리 위축을 야기하면서

소프트웨어개발완료보고서

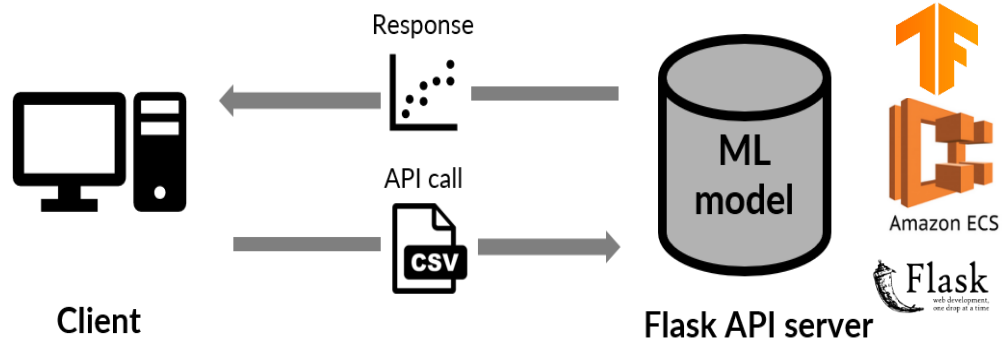
출산을 감소, 빈부격차 확대와 같은 여러 사회문제의 원인으로 작용하고 있다.[1] 지방에서의 주거불안정 문제는 특히 젊은 세대 인구유출의 원인이 되는 민감한 사안이다. 주택가격의 급등의 원인으로서는 수요-공급의 불균형, 특정지역 투기과열 등 여러가지가 복합적인 원인이 존재한다. 이에 정부에서는 주택 공급량 증가, 투기과열지구 지정 등 여러가지 규제와 정책을 시행하여 상승률 감소에 나서고 있다. 최근 경남에서는 창원시 의창구가 투기과열지구로 지정되면서 주택법에 따른 분양권 전매제한, 주택청약 1순위 자격제한, 주택담보대출 조건강화 등의 규제가 적용되었다.[2] 하지만 규제가 과도해지면 이는 또한 주택가격의 폭락으로 이어질 수 있다. 주택가격이 폭락하게 되면 부동산 시장이 위축되면서 건설업체들과 부동산으로 자산을 보유하고 있는 중산층이 피해를 입고 가격 폭락 전 부동산을 담보로 금융기관으로 대출을 받은 개인 또는 기업들이 부동산을 팔아서 채무를 이행할 능력이 떨어지는데 이런 상황이 지속되면 경기침체로 이어질 수 있다[3]. 따라서 주택가격의 등락폭을 최소화하면서 적절한 가격을 유지하는 것이 시장경제를 유지하는데 있어서 중요한 요소인데, 이는 정부 및 지자체에 요구되는 중요한 능력이자 조직의 역량을 평가하는 지표가 되기도 한다[4]. 따라서 적절한 주택가격 유지를 위해서는 철저한 시장조사를 통해 현재 경제상황을 파악하고 여러가지 경제지표를 객관적으로 분석하여 예상되는 주택가격지수를 산정하여 이에 맞는 적절한 규제나 정책을 시행해야 한다. 이때, 정책수립 시 미래 주택가격지수를 예측할 수 있는 딥러닝 모델의 예측값을 참고한다면 객관적이고 정확도가 높은 예측값을 산정할 수 있을 것으로 생각되며 주택정책 수립에 필요한 신뢰성 있는 참고자료가 될 수 있을 것으로 생각된다.

4.1 목표

창원시 주택가격지수와 관련성이 있는 경제지표 데이터를 학습한 모델을 사용하여 딥러닝 관련 지식이 없는 사람들도 딥러닝 모델을 통해 주택가격지수를 예측 값을 제공받을 수 있는 프로그램을 제작하는 것이 목표이다. 경제지표 데이터를 월별 시계열 데이터셋으로 확보하고 예측모델은 시계열데이터 학습에 적합하다고 알려진 LSTM 모델을 사용한다.[5][6][7] 창원시에 소속된 각 지역구 주택가격지수 예측에 최적화된 모델이 존재하고 사용자는 주택가격지수를 예측할 지역구를 선택할 수 있다, 예측모델은 미리 가장 높은 예측성능을 보이는 하이퍼 파라미터를 선정하여 튜닝한 모델을 사용하여 예측 프로그램을 제작한다. 최종적으로는 GUI 환경에서 경제지표 데이터를 입력하면 그래프와 함께 예측치를 출력할 수 있도록 구현하고자 한다.

5. 시스템 명세

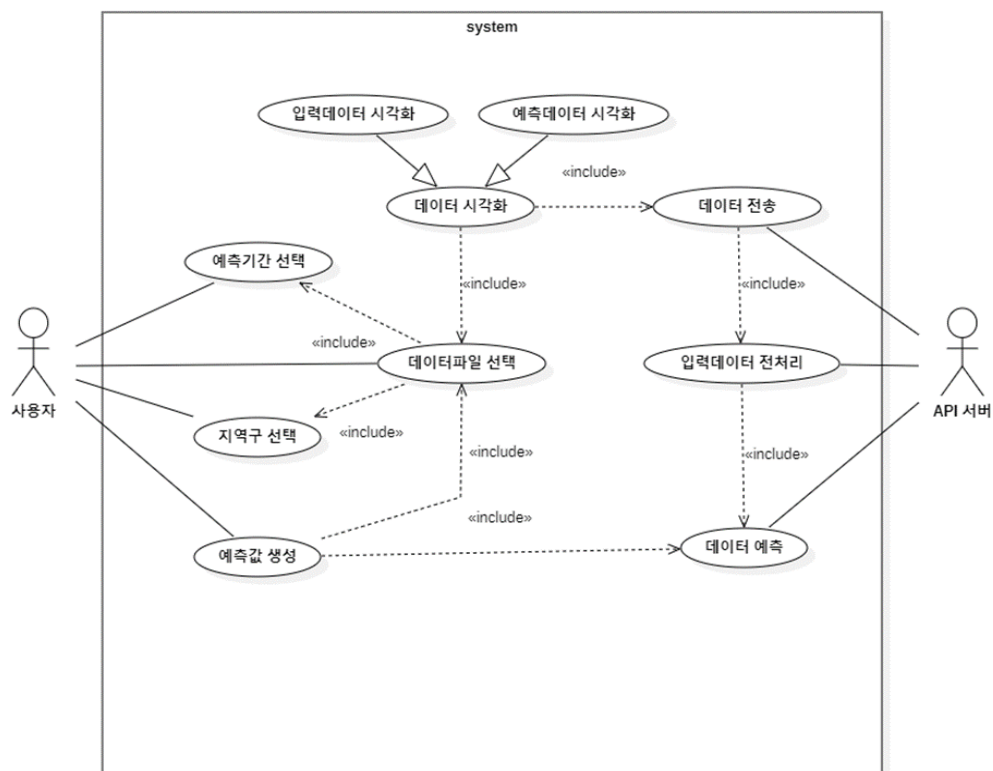
5.1 구성



[그림 1. 시스템 구조도]

시스템은 데이터셋을 API call 형태로 전달하는 클라이언트와 요청받은 데이터를 전처리하여 학습모델의 입력으로 전달하고 출력된 예측값을 응답으로 클라이언트에 반환하는 API서버로 구성되어 있다. 클라이언트는 GUI를 사용하여 예측하고자 하는 지역과 기간을 선택할 수 있으며, 데이터파일을 선택 후 POST API 방식으로 서버에 전달한다. 서버는 받은 데이터를 내부에서 전처리하고 저장된 학습모델을 로드하여 전처리한 데이터를 입력으로 예측값을 생성한다. 예측 값은 실제데이터 형식으로 다시 변환하여 클라이언트에 응답으로 문자열형태로 반환된다. 반환된 예측 데이터는 GUI를 통해 그래프로 시각화 되어 클라이언트에게 제공된다.

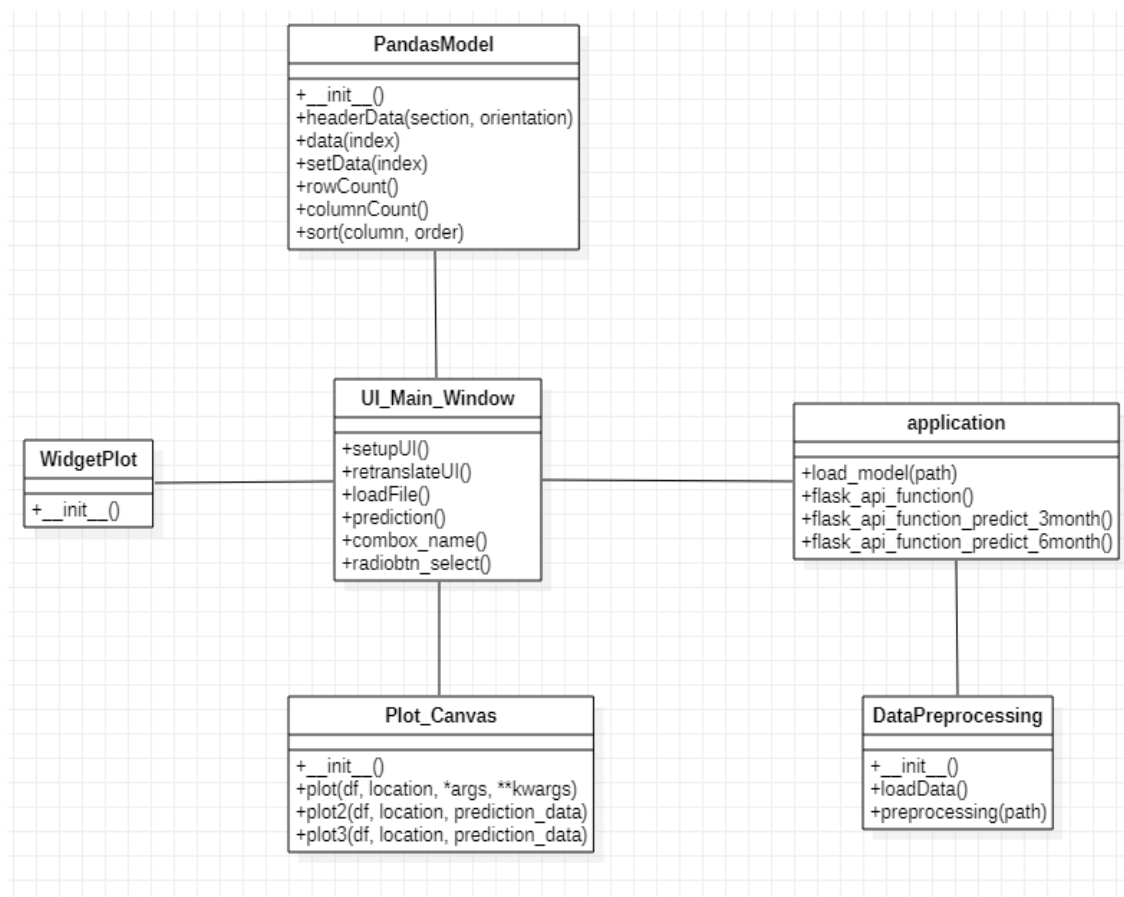
5.2 USE-CASE 다이어그램



[그림2. 시스템 Use-Case diagram]

그림2에서사용자는 GUI를 통해 미래 주택가격지수를 예측할 기간(3개월, 6개월)을 선택할 수 있다. 사용자가 예측 지역구를 선택하게 되면 해당지역에 대한 데이터를 프로그램에 업로드 할 수 있다. 업로드된 데이터는 그래프로 시각화되어 사용자에게 제공되고, 데이터 값을 테이블 형태로 열람할 수 있다. 데이터가 업로드 되면 예측값을 생성할수 있다. ‘예측값 생성’을 선택하면 업로드된 데이터가 미리 학습된 모델이 저장되어 있는 서버로 전송된다. 전송된 데이터는 전처리 후 학습모델을 통과하여 주택가격지수 예측값을 생성한다. 생성된 예측값은 다시 사용자 프로그램으로 전송되어 사용자에게 시각화된 예측값 데이터를 제공한다.

5.3 CLASS 다이어그램



[그림3. 시스템 Class diagram]

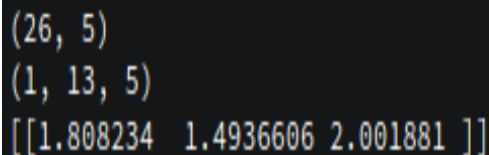
클래스는 6개로 구성되어 있으며, 그림3에서 사용자의 로컬 환경에서 GUI를 제공하고 API 서버와 통신하는 클래스는 UI_MainWindow클래스이다. 데이터 시각화는 WidgetPlot, PlotCanvas 클래스가 담당하고 matplotlib 라이브러리를 사용한다. PandasModel 클래스는

소프트웨어개발완료보고서

사용자가 업로드한 데이터파일을 데이터 처리 라이브러리인 Pandas를 사용하여 학습모델의 입력 데이터 형식에 맞게 데이터를 조작하는 클래스이다. Application 클래스는 API 서버에서 동작하는 클래스로 UI_MainWindow 클래스에서 prediction() 클래스가 Post API 방식으로 데이터와 함께 요청을 전송하면 데이터는 DataPreprocesssing 클래스에서 Tensor 형태로 전처리되어 서버상의 학습모델 바이너리로 전달된다. 학습모델이 예측값을 생성하면 예측 데이터는 다시 사용자 로컬 환경의 UI_MainWindow 클래스로 반환된다. 반환된 값은 Plot_Canvas 클래스로 전달되어 예측값에 대한 시각화 그래프를 생성하여 사용자에게 제공한다.

5.4 API 서버

프로그램 동작을 위한 설치 의존성과 호환성 문제를 줄이기 위해 딥러닝 모델 실행에 있어 용량이 크고 설치가 까다로운 텐서플로우와 CUDA, cuDNN가 실행환경에 세팅되어 있지 않더라도 학습모델로부터 계산된 출력결과를 얻을 수 있도록 미리 학습된 모델을 로컬 환경에서 분리하고 WEB API 환경을 구성하였다. 프로그램에서는 URL을 통해 서버로 데이터를 전송하면 서버로 전송된 데이터가 학습모델을 통과하여 예측치를 응답으로 전송하도록 구현했다. 따라서 사용자는 라이브러리나 운영체제의 의존성과 관계없이 인터넷만 연결되어 있으면 프로그램을 실행할 수 있다. WEB API는 python 기반의 Flask Framework[8]를 사용하였다. 웹서버는 goorm[9]에서 Aws기반 컨테이너를 외부접속 포트가 개방된 웹서버 환경으로 구현하였다. 웹서버 OS는 Linux기반 UBUNTU배포판을 사용한다. 활성화된 웹서버에 미리 학습된 모델 바이너리파일을 업로드하고 데이터 전처리, 데이터 예측, 2가지로 나누어 API URL을 구성했다. 데이터 예측의 경우 예측기간과 지역에 따라 사용되는 학습모델이 다르므로 지역구와 예측기간의 조합별로 URL을 따로 구성했다. 실제 프로그램으로부터 API call을 통해 업로드된 데이터가 모델을 통과한 3개월 예측값은 그림4와 같다. 예측값은 문자열 형태로 로컬 환경의 프로그램으로 전달되어 형변환 과정을 거쳐 그래프로 시각화된다.



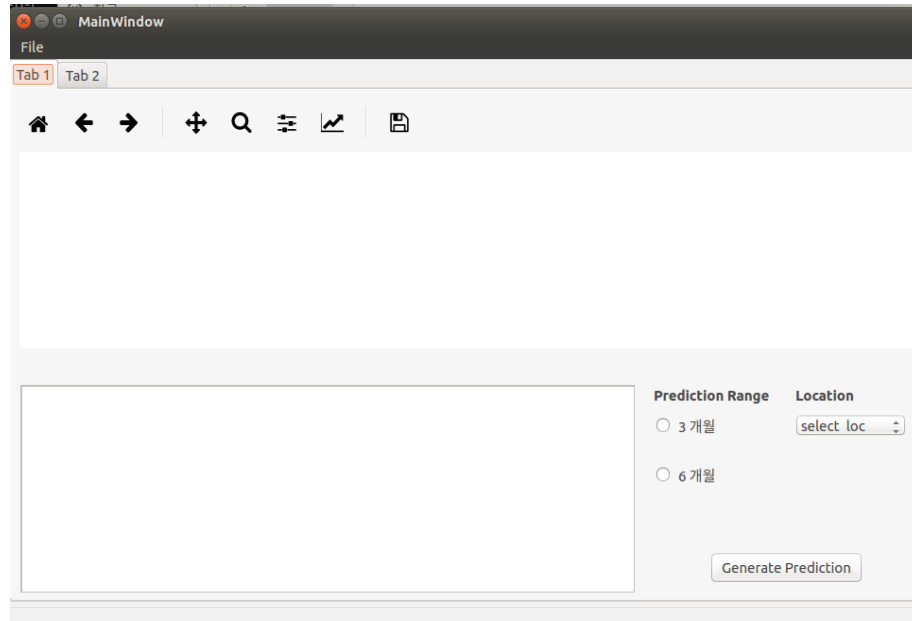
```
(26, 5)
(1, 13, 5)
[[1.808234 1.4936606 2.001881 ]]
```

[그림4. 학습모델이 예측값이 출력된 웹 서버 Terminal 캡처]

5.5 GUI

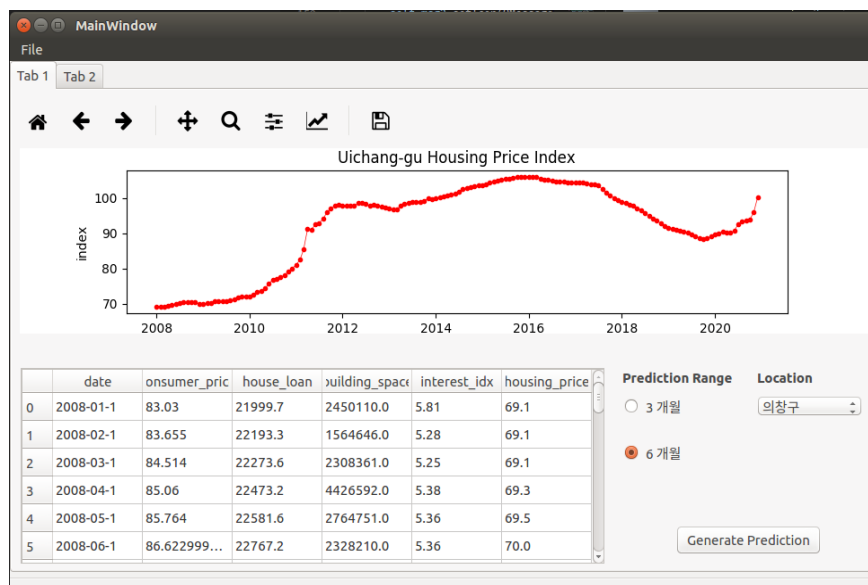
소프트웨어개발완료보고서

GUI는 python 기반 GUI 제작도구인 pyqt5[10]로 제작하였다. 라디오 버튼으로 예측기간을 선택하고 콤보박스를 통해 예측하고자 하는 창원시 지역구를 선택할 수 있다. 좌측상단 탭 버튼으로 예측을 위한 지역구 데이터를 GUI기반으로 선택하여 프로그램으로 업로드 하게되면 자동으로 그래프 형태로 시각화되어 사용자에게 제공된다. 또한 컬럼에



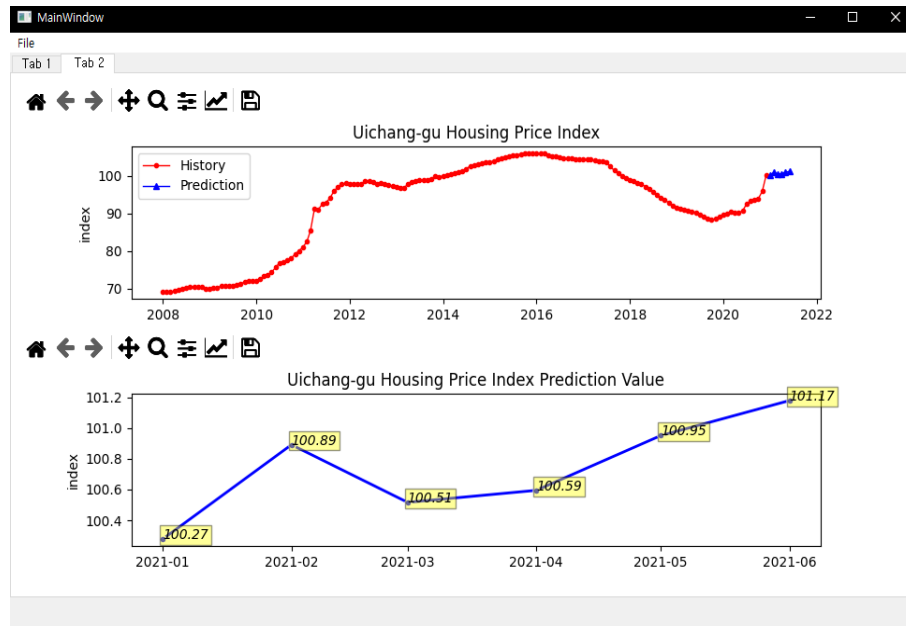
[그림5. 프로그램 초기실행화면]

따라 자동으로 데이터 테이블을 생성하여 기간에 따라 세부 값을 열람할 수 있다. 그래프 시각화는 matplotlib[11] 라이브러리를 사용하여 구현하였으며 그래프 축 단위 변경, 이동, 확대, 축소와 png 형식의 이미지 형태로 저장을 지원한다. 또한 커서를 통해 그래프 상의 각 점들의 값을 확인할 수 있다. 데이터파일을 프로그램에 업로드 했을 때 시각화된 그래프는 그림6 과 같다. 이후 ‘Generate Prediction’ 버튼을 클릭하면 탭이



[그림6. 데이터 시각화 GUI]

자동으로 전환되고 예측값을 시각화하여 사용자에게 제공하는데, 기존 데이터 분포에 예측값을 더한 그래프와 예측값만을 월단위로 값 레이블과 함께 세부적으로 표현한 그래프로 구성된다. 전환된 탭의 그림7과 같다.



[그림7. 예측값 데이터 시각화 GUI]

6 학습모델 명세

6.1 데이터 선정

주택매매가격지수

$$I_j = \prod_{i=1}^{n_j} (x_i)^{1/n_j}$$

단,

I_j : j 지역 지수

$$x_i = \frac{P_i^t}{P_i^0}$$

P_i^0 : i 표본의 기준시점 조사가격

P_i^t : i 표본의 t 시점(비교시점) 조사가격

n_j : j 지역 표본 수

[그림8. 주택매매가격지수 산정공식]

- 전국 주택시장의 매매, 전세, 월세 가격을 조사하여 주택시장의 평균적인 가격 변화를 측정, 주택시장 판단지표 또는 주택정책수립의 참고자료 제공하기 위해

소프트웨어개발완료보고서

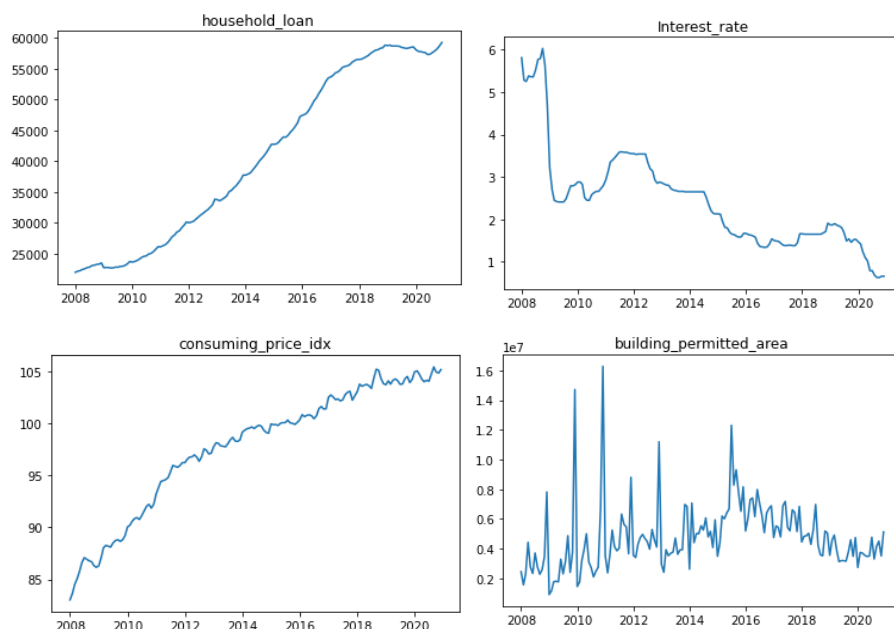
주택법 제88조, 제89조 및 동법 시행령 제91조에 의거하여 한국부동산원에서 매월 시군구별로 발표하는 통계이다. 전국 261개 시군구의 거래 가능한 아파트, 단독주택, 연립주택이 조사대상에 포함된다. 편제방법으로는 기준시점 대비 조사시점 가격비의 기하평균방식으로 지수를 산출한다. Jevons index 방법론 적용하여 지수산정 공식은 아래와 같다. 한국부동산원 통계정보시스템[12]에서 2003년 12월부터 현재까지 시군구별 데이터를 제공한다.

- 독립변수 데이터

- 주택가격지수 예측에 대한 기존 선행연구[5][6][7][13] 들을 참고하여 주택가격과 연관성이 있다고 알려진 데이터셋 5가지를 선정하였다.

데이터 명	정보	단위
금리	월별 전국 시장금리.	(%)
가계대출금	월별 순수가계에 대한 금융기관 대출 포괄.	억 원
소비자물가지수	월별 시군구별 소비자물가지수	(%)
건축허가면적	건축행정정보시스템(AIS)과 행정문서를 통해 시군구 건축 인.허가 통계	m ³
주택가격지수	-	-

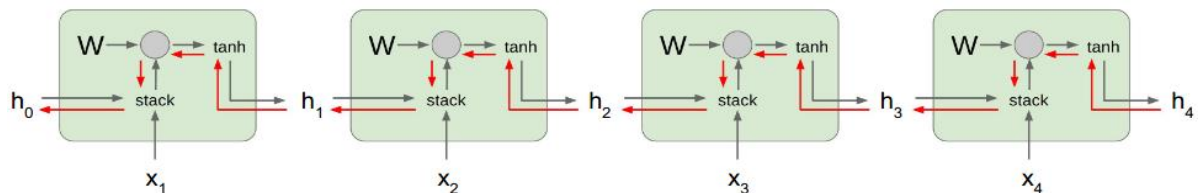
6.2 데이터 수집



[그림9. 수집 데이터 시각화]

각 독립변수, 종속변수 데이터는 한국 부동산원과 한국은행 통계정보 시스템[14]에서 수집하였다. 2008년 1월1일부터 2020년 12월 1일까지 월별 데이터를 수집하였으며 각 데이터를 시각화한 그래프는 그림9과 같다. 경제지표 데이터 특성상 시간의 흐름에 따른 추세가 관찰되는데 전통적인 시계열 예측모델의 경우에는 예측성능을 높이기 위해서 시간이 변해도 통계적 특성이 일치하는 정상성을 가진 데이터가 요구된다. 따라서 데이터의 예측을 위해서는 데이터의 정상성을 확보하기 위해 연이은 값들 간의 차이를 계산하여 그래프를 변환하는 차분(differencing)과정이 필요하다. 그러나 딥러닝 모델의 경우에는 신경망의 특성상 데이터의 정상성을 고려할 필요가 없고 예측성능도 기존 예측모델보다 뛰어나다는 장점이 있다.

6.3 학습모델 선정



[그림 10. RNN모델의 구조]

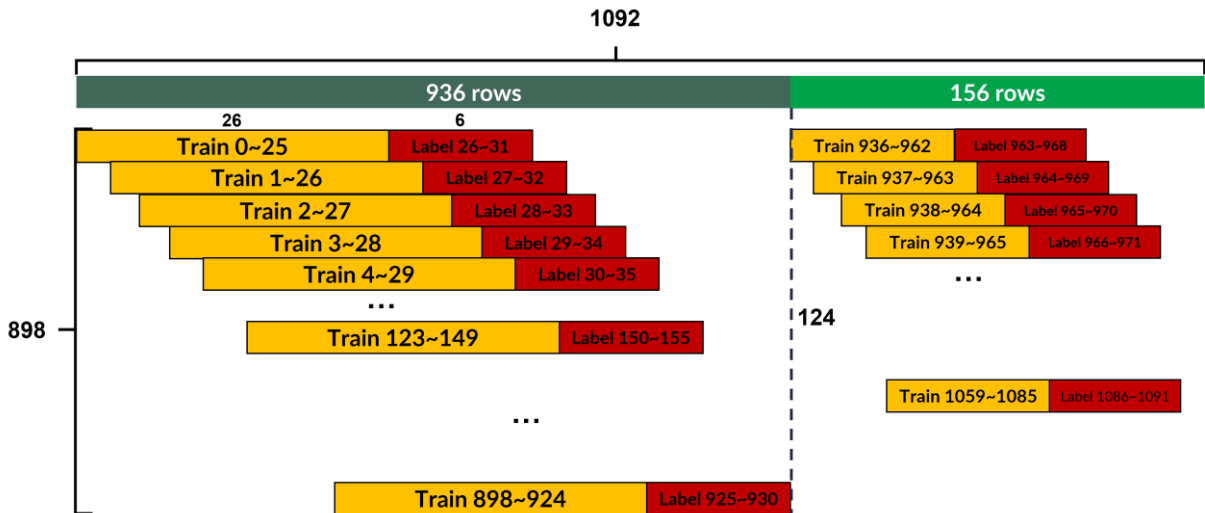
주택가격예측을 위해 다변량 시계열 데이터를 학습하는 LSTM[15] 모델을 사용한다. LSTM(Long Short-Term Memory)은 순환신경망(Recurrent Neural Network)의 한 종류로, 은닉층이 순환구조로 구성되어 각 레이어마다 이전 레이어의 hidden state(h_{n-1}) 값을 반영하여 현재 입력값과 가중치를 계산한 hidden state(h_n)를 다음 레이어로 전달하여 가중치 계산에 반영한다. 즉, 과거의 데이터가 미래에 영향을 끼치는 구조를 가진 모델이다. RNN은 음성인식, 번역, 감정분석, 자연어 처리 등에서 높은 성능을 보이며 또한 시계열데이터 학습에도 적합하다고 알려져 있다. 학습모델은 머신러닝 Framework인 Tensorflow[16][17]를 사용하여 구현하였다.

6.4 데이터 전처리

일반적으로 딥러닝 모델 훈련을 위해서는 원본 데이터를 신경망이 GPU에서 학습할 수 있도록 고차원 행렬 형태인 Tensor형식으로 변환해야 한다. 이러한 과정을 데이터 전처리라고 하는데, 이 과정에서 학습모델의 훈련성능을 높이기 위한 여러가지 기법들이 존재한다. 여기서는 기존에 확보한 데이터의 양이 모델이 학습하기에 충분하기 않기 때문에 동일한 데이터를 중첩시켜 기존보다 많은 양의 데이터를 학습하는 효과를 얻을 수 있는 Sliding window 기법을 사용했다. 아래 그림11과 같이 전체 데이터가 1092행이고 이중 훈련데이터를 936행으로 분리했을 때, 학습모델은 훈련데이터에서 한

소프트웨어개발완료보고서

번에 26개월(26행) 데이터를 학습하고 6개의 예측값을 출력하여 Label(정답)데이터와의 오차를 계산하여 가중치를 반영한다고 할 때, Sliding window 기법을 사용하면 다음 학습 데이터는 이전 학습데이터(window)에서 1칸씩만 이동(sliding)하여 훈련을

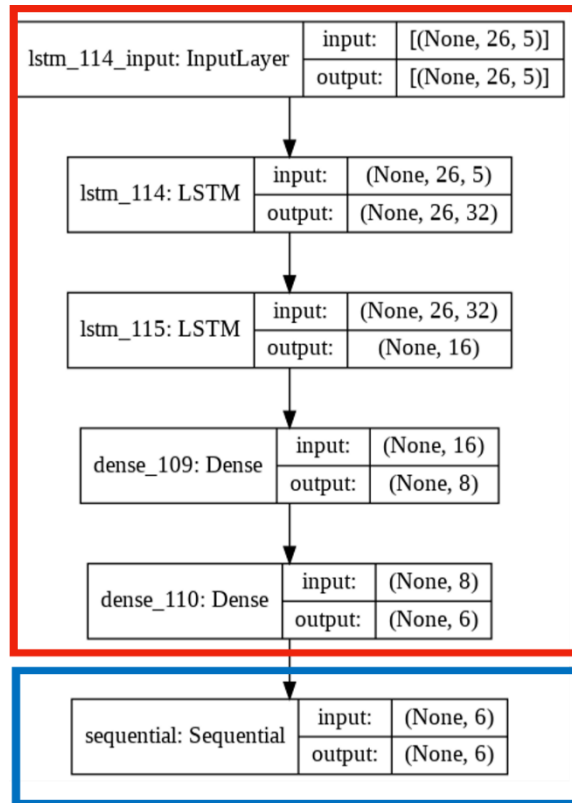


[그림11. Sliding window]

진행한다. 이는 실제 데이터 보다 더 많은 데이터를 학습하는 것과 비슷한 효과를 가진다. 원본데이터는 데이터 분석 라이브러리인 pandas를 사용하여 로드한 뒤 고차원 행렬로 변환하고 sliding window 형태 데이터 구조를 변경한다.

6.5 전이학습과 미세조정을 사용한 모델 성능 최적화

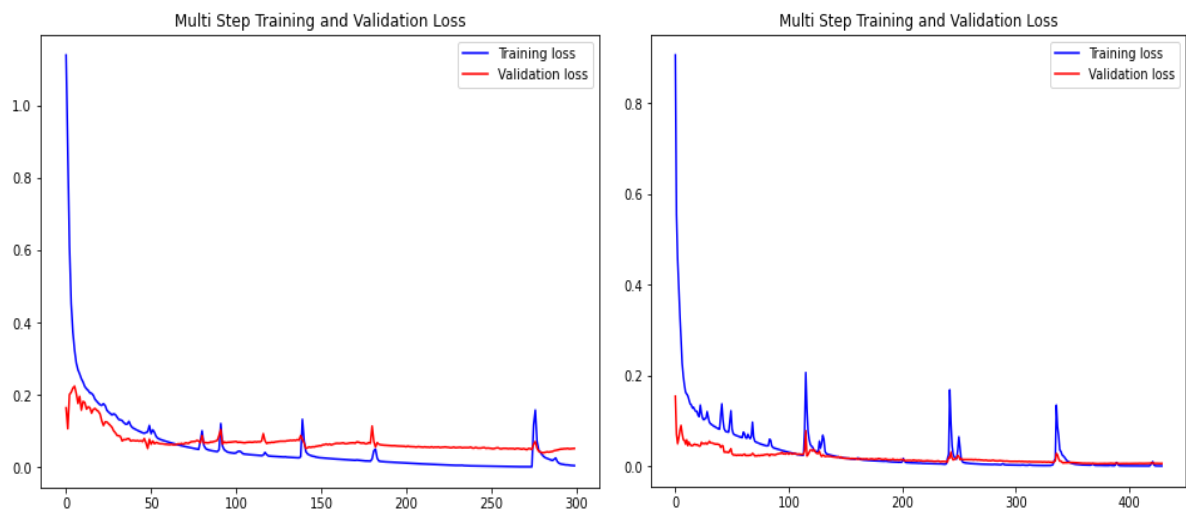
딥러닝에서 전이학습(Transfer learning)[18]이란 상대적으로 많은 양의 데이터로 미리 학습된 모델을 사용하여 비슷하거나 전혀 새로운 데이터를 학습하는 모델에 사용하는 기법이다. 이미지 분류모델을 제작할 때 성능이 검증된 이미지분류모델인 Resnet[19], ImageNet[20], VGG[21]모델을 가져와 사용자가 분류하고자 하는 이미지데이터로 추가적으로 학습시키거나 신경망을 추가하는 것을 전이학습의 한 예라고 말할 수 있다. 미세조정(Fine tuning)은 미리 학습된 학습모델 신경망 가중치를 새로운 데이터 예측에 적합하도록 업데이트하여 기존학습모델의 성능을 유지하면서 새로운 데이터 특성에 맞게 가중치 변동량을 줄여 세밀하게 가중치를 튜닝하는 것을 말한다. 전이학습은 주로 학습데이터의 양이 작은데 높은 예측성능이 요구될 경우 사용한다. 창원시 주택가격지수 예측모델은 학습데이터의 양이 156행으로 sliding window를 구성해도 LSTM 모델 학습에 있어 데이터의 양이 절대적으로 부족하다. 따라서 전이학습 기법을 도입하여 데이터 양상이 비슷하고 비교적 확보가 용이한 각 광역시별 주택가격지수 데이터를 구성하여 사전학습모델을 제작했다. 확보한 광역시별 전체 데이터셋이 1920 line일 때 sliding window를 구성하면 학습데이터로 23,348 line(26 x 898)을 모델이 학습하는 효과를 볼 수있다. 전이학습 기법을 사용하여 학습모델을 구성했을 때 모델의 구조는 아래와 같다.



[그림 12. 학습모델구조]

붉은색 테두리 부분의 모델 레이어는 광역시별 데이터로 사전학습을 진행한 모델이다. 해당 모델은 데이터 학습 후 가중치 반영률을 낮게 조정하여 기존 예측성능을 유지할 수 있도록 구성했다. 그리고 1개의 새로운 단순연결 신경망을 추가하여 창원시 데이터에 대해서 가중치를 강하게 반영할 수 있도록 하였다.

6.6 예측성능

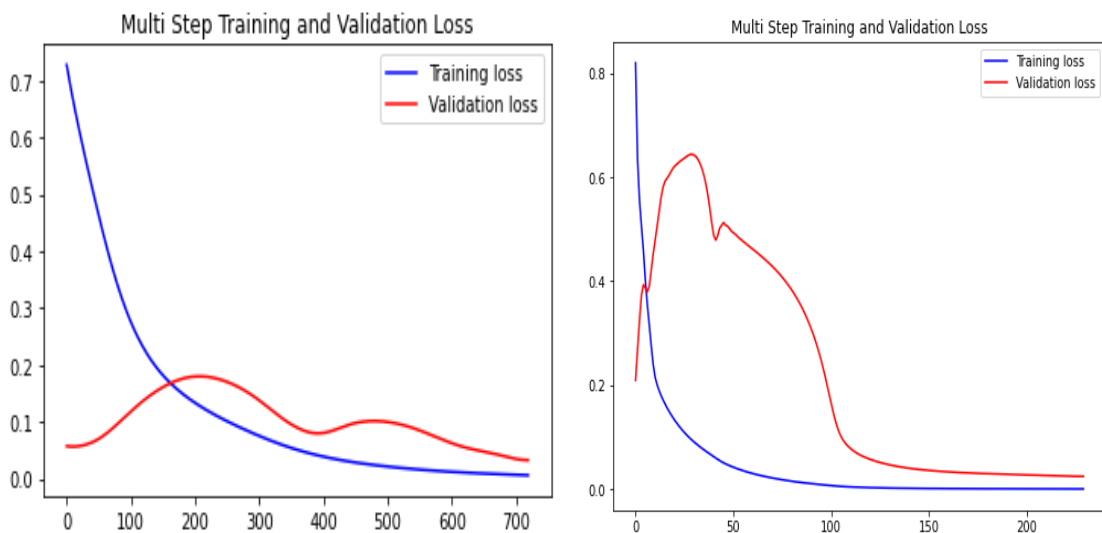


[그림 13. 사전학습모델 Loss]

주택가격지수 예측모델의 예측성능을 확인하기 위해 Loss를 측정하고 실측데이터와 비교를 진행했다. Loss는 데이터를 학습한 모델이 출력한 값과 정답 데이터와 오차값을 의미하며 0에 가까울수록 예측성능이 높다고 판단한다. 광역시 데이터를 학습한 사전학습모델의 Loss 그래프는 그림13과 같다. Training loss는 훈련데이터에서의 정답데이터와 예측데이터의 Loss이고 Test loss는 학습모델 성능 확인을 위해 훈련데이터로 사용하지 않은 검증데이터와 예측데이터의 loss이다. 왼쪽 그래프는 3개월 예측모델에 대한 Loss 분포이고 오른쪽 모델은 6개월 예측에 대한 Loss 분포이다. 그래프에서 x축은 epoch수를 의미하는 것으로 딥러닝에서 epoch은 전체 훈련데이터에 대해 1회 학습을 완료했을 때 1epoch이라고 한다 따라서 3개월 예측 모델의 경우 epoch 300 이상부터 Loss가 점점 증가하는 추세를 보이고 6개월 예측모델의 경우 epoch400 이후 0에 거의 수렴하는 Loss값을 보인다. 사전학습모델의 변수설정과 마지막 epoch에서의 Loss값은 아래 표와 같다.

	6month model	3month model
Batch_size	512	256
Buffer_size	4096	4096
Evaluation_Interval	28	28
Optimizer	Adam	Adam
Loss	MSE	MSE
Validation_step	2	2
Epoch	300	430
Validation Loss	0.0523	0.0062

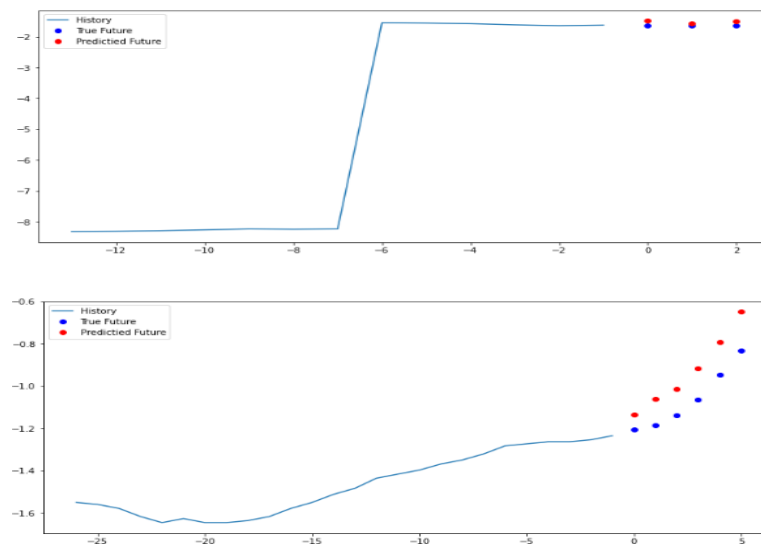
그림14는 사전학습모델에 1개 레이어를 추가하고 창원시 데이터로 미세조정을 통한 전이학습을 진행한 학습모델의 Loss 값이다.



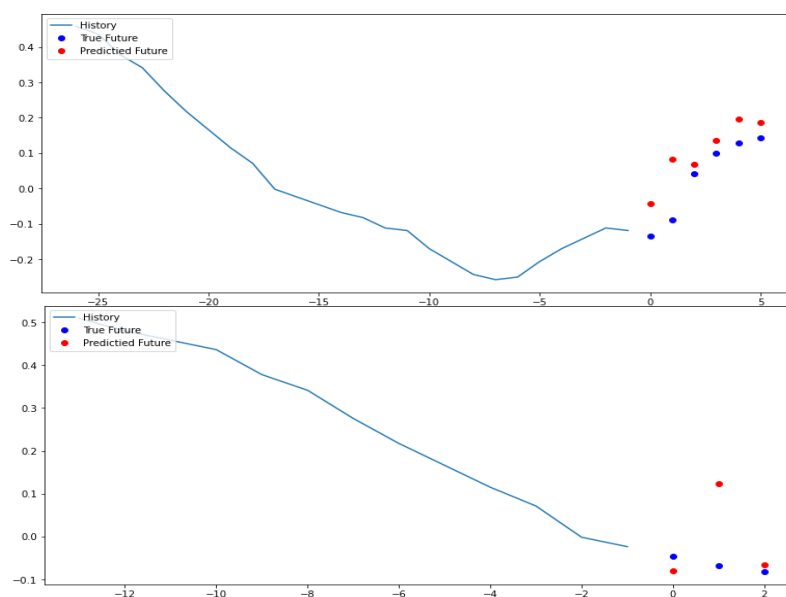
[그림 14. 전이학습모델 Loss]

소프트웨어개발완료보고서

왼쪽이 3개월 예측 전이학습 모델이고 오른쪽이 6개월 예측 전이학습 모델이다. 전이학습을 진행하기 전 Loss 그래프보다 Loss 값이 증가한 것을 확인할 수 있는데, 이는 학습데이터 양이 작아 모델이 비교적 충분히 창원시 데이터를 학습하지 못한 것으로 추측된다. 이는 차후 창원시 데이터가 어느정도 충분히 축적되면 더 높은 성능을 보일 것으로 기대된다. 사전학습모델과 전이학습모델의 실측값 비교 그래프는 그림15, 16에서 확인할 수 있다. 예측해야할 데이터가 많은 6개월 예측모델의 예측값 보다 3개월 예측모델의 예측값이 실측값과 더 비슷한 것을 확인할 수 있다. 아래와 같다. 사전학습모델의 실측결과와 달리 비교적 예측데이터에 노이즈가 발생한 것을 확인할 수 있다.



[그림 15. 사전학습모델 실측값]



[그림 16. 전이학습모델 실측값]

7 참고문헌

- [1]. 박진백, 이재희, "경기변동에 따른 주택가격변동이 출산율에 미치는 영향", 2016, 육아정책연구소, 2016, vol.10, no.3, 통권 20호 pp. 51-69 (19 pages)
- [2]. 국토교통부, "창원 의창구 투기과열지구...부산·대구 등 36곳 조정대상지역 지정", 대한민국 정책브리핑, 2020.12.17, <https://www.korea.kr/news/policyNewsView.do?newsId=148881292>
- [3]. 주 OECD 대한민국 대표부, "주택시장과 거시경제에 관한 연구 결과 및 시사점", 2019.8.30, https://overseas.mofa.go.kr/oecd-ko/brd/m_20806/view.do?seq=112
- [4]. 최차순, "부동산정책이 부동산시장에 미치는 영향에 관한 연구", 2010, 대한부동산학회지, vol. 28, no. 2, pp. 69-91.
- [5]. 이태형, 전명진, 2018, "딥러닝 모델을 활용한 서울 주택가격지수 예측에 관한 연구", 주택도시연구, 8(2), 39-56.
- [6]. 전해정, 양혜선. 2019, "딥 러닝을 이용한 주택가격 예측에 관한 연구", 주거환경, 17(2), 37-49.
- [7]. 민성욱, 2017, "딥러닝(Deep Learning)을 이용한 주택가격 예측모형 연구", 강남대학교 대학원 박사학위 논문
- [8]. "flask reference", flask, last modified 2021.5.10, accessed 2021.5.14, <https://flask-docs-kr.readthedocs.io/ko/latest/>
- [9]. Goorm ide platform, <https://ide.goorm.io/>
- [10]. PyQt5 Reference, <https://www.riverbankcomputing.com/static/Docs/PyQt5/introduction.html#pyqt5-components>
- [11]. Matplot document, <https://matplotlib.org/>
- [12]. "한국부동산원", <http://www.kab.co.kr/> 24.
- [13]. 정원구·이상엽, 2007, "인공신경망을 이용한 공동주택 가격지수 예측에 관한 연구 -서울지역을 중심으로-", 주택연구, 15(3), pp.39~64.
- [14]. "한국은행 경제통계시스템", <http://ecos.bok.or.kr/>
- [15]. Fei-Fei Li, Justin Johnson, 2017, "Lecture 10: Recurrent Neural Networks", <http://cs231n.stanford.edu/slides/2017/>
- [16]. "Complete guide to transfer learning & fine-tuning in Keras", Keras Guide, last modified 2021.5.12, accessed 2021.5.14, https://keras.io/guides/transfer_learning/
- [17]. Tensorflow Reference, <https://www.tensorflow.org/versions?hl=ko>
- [18]. Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359.
- [19]. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings

소프트웨어개발완료보고서

of the IEEE conference on computer vision and pattern recognition. 2016.

[20]. J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[21]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

8 부록

본 프로젝트 수행 시에 각 단계별 산출물에 대한 제출 문서는 아래와 같음

NO	산출물명	수량	매체	제출시기	비고
1	개발계획서	1	문서	시작 후 2주 이내	
2	분석/설계명세서 USECASE 다이어그램(요구사항분석서) 기능차트 아키텍처 정의서 용어사전 클래스다이어그램(ERD), 시퀀스다이어그램(DFD), 테이블목록 테이블정의서 프로그램목록 User Interface 설계서		문서	* ERD는 2021.10.15일 필수제출이며 기타자료도 제출 가능함.	
5	사용자 메뉴얼	1	문서	2021.10.15	
6	운영 메뉴얼	1	문서	2021.10.15	
7	프로젝트 완료보고서, 구동경로, 시연 동영상 등 인수인계 확인서(제출로 대체함)	1	문서	2021.10.15	

※ 산출물의 소요 발생시 수량 추가 및 수량 변경은 조정가능.