



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2022년 8월
석사학위 논문

딥러닝을 활용한
농산물 가격예측 모델 설계

Design of Price Prediction Model of
Agricultural Product Using Deep Learning

광주대학교
GWANGJU UNIVERSITY

광주대학교 대학원

정보통신공학과

이 현 석

딥러닝을 활용한 농산물 가격예측 모델 설계

Design of Price Prediction Model of
Agricultural Product Using Deep Learning



2022년 06월 13일

광주대학교 대학원

정보통신공학과

이 현 석

딥러닝을 활용한 농산물 가격예측 모델 설계

지도교수 김 광 현

이 논문을 공학 석사학위신청 논문으로 제출함.

2022년 06월 13일

광주대학교 대학원

정보통신공학과

이 현 석

이현석의 석사학위논문을 인준함

위원장	광주대학교 교수	조 정 호 (인)
위 원	광주대학교 교수	이 유 (인)
위 원	광주대학교 교수	김 광 현 (인)

2022년 06월 13일

광주대학교
광주대학교 대학원
GWANGJU UNIVERSITY

목 차

ABSTRACT

제1장 서론	1
제1절 연구 배경	1
제2절 연구 방법	2
제2장 관련 연구	4
제1절 기존 농산물 가격예측.....	4
제2절 농산물 가격예측을 위한 딥러닝 기법	7
1. RNN(Recurrent Neural Network)	7
2. LSTM(Long Short Term Memory)	8
3. GRU(Gated Recurrent Units)	10
제3장 딥러닝을 활용한 농산물 가격예측 모델 연구 ·	12
제1절 실험 데이터	12
1. 데이터 수집	12
2. 데이터 분석	14
3. 데이터 전처리	19
제2절 딥러닝 모델.....	26
1. RNN 예측모델	26
2. LSTM 예측모델	32

3. GRU 예측모델	38
제4장 실험 평가 방법 및 결과	45
제1절 실험 평가 방법	45
제2절 실험 결과	46
제5장 결론 및 향후 연구	48
참고 문헌	49



그림 목차

[그림 2-1] RNN 구조	7
[그림 2-2] LSTM 내부구조	9
[그림 2-3] GRU 내부구조	10
[그림 2-4] 반복횟수, 계산시간을 고려한 학습 곡선	11
[그림 3-1] 기상청 날씨 데이터 - 평균기온	15
[그림 3-2] 기상청 날씨 데이터 - 평균 상대습도	15
[그림 3-3] 기상청 날씨 데이터 - 평균 풍속	16
[그림 3-4] 기상청 날씨 데이터 - 60일간 평균 풍속	16
[그림 3-5] 소비자 물가 데이터	17
[그림 3-6] 농산물 양과 가격 데이터	17
[그림 3-7] 농산물 무 가격 데이터	18
[그림 3-8] 농산물 배추 가격 데이터	18
[그림 3-9] 기상청 날씨 데이터	19
[그림 3-10] (좌)월별 소비자 물가 데이터, (우)일별 소비자 물가 데이터	20
[그림 3-11] 농산물 가격예측을 위한 데이터 셋	21
[그림 3-12] 양과 가격예측을 위한 데이터 셋	21
[그림 3-13] 평균기온(°C) train, test 데이터	22
[그림 3-14] 평균 풍속(m/s) train, test 데이터	22
[그림 3-15] 평균 상대습도(%) train, test 데이터	22
[그림 3-16] 소비자 물가 train, test 데이터	23
[그림 3-17] 농산물 양과 가격 train, test 데이터	23
[그림 3-18] 농산물 무 가격 train, test 데이터	23
[그림 3-19] 농산물 배추 가격 train, test 데이터	24
[그림 3-20] 양과 데이터 셋 스케일링	24
[그림 3-21] 양과 데이터 셋 전처리 후	25
[그림 3-22] RNN 모델 양과 가격예측 결과	31
[그림 3-23] RNN 모델 무 가격예측 결과	31

[그림 3-24] RNN 모델 배추 가격예측 결과	32
[그림 3-25] LSTM 모델 양파 가격예측 결과	37
[그림 3-26] LSTM 모델 무 가격예측 결과	37
[그림 3-27] LSTM 모델 배추 가격예측 결과	38
[그림 3-28] GRU 모델 양파 가격예측 결과	43
[그림 3-29] GRU 모델 무 가격예측 결과	43
[그림 3-30] GRU 모델 배추 가격예측 결과	44
[그림 4-1] 농산물 가격예측 모델별 RMSE 그래프	46
[그림 4-2] 농산물 가격예측 모델별 정확도 그래프	47



표 목차

[표 2-1] Lasso 회귀분석을 활용한 농산물 가격예측 모델 연구 결과	4
[표 2-2] LSTM 네트워크를 활용한 농산물 가격예측 모델 정확도	5
[표 2-3] 인공신경망 기법을 이용한 최적의 농산물 가격예측모델 정확도	6
[표 3-1] 기상청 날씨 데이터	12
[표 3-3] 양파, 무, 배추 가격 데이터	13
[표 3-4] RNN 모델 구성도	26
[표 3-5] RNN 모델 하이퍼 파라미터	27
[표 3-6] 양파, 무, 배추 RNN 모델 오차	28
[표 3-7] RNN 모델 양파 가격예측 결과	28
[표 3-8] RNN 모델 무 가격예측 결과	29
[표 3-9] RNN 모델 배추 가격예측 결과	30
[표 3-10] LSTM 모델 구성도	32
[표 3-11] LSTM 모델 하이퍼 파라미터	33
[표 3-12] 양파, 무, 배추 LSTM 모델 오차	33
[표 3-13] LSTM 모델 양파 가격예측 결과	34
[표 3-14] LSTM 모델 무 가격예측 결과	35
[표 3-15] LSTM 모델 배추 가격예측 결과	36
[표 3-16] GRU 모델 구성도	38
[표 3-17] GRU 모델 하이퍼 파라미터	39
[표 3-18] 양파, 무, 배추 GRU 모델 오차	39
[표 3-19] GRU 모델 양파 가격예측 결과	40
[표 3-20] GRU 모델 무 가격예측 결과	41
[표 3-21] GRU 모델 배추 가격예측 결과	42
[표 4-1] 농산물 가격예측 모델별 RMSE	46
[표 4-2] 농산물 가격예측 모델별 정확도	47

ABSTRACTS

Design of Price Prediction Model of Agricultural Product Using Deep Learning

Lee Hyun Suk

Advisor : Prof. Kim Gwang Hyun, Ph. D.

Department of Information and Communication Engineering

Graduate School of Gwangju university

We do not have a fixed price for agricultural products sold domestically, and the price of agricultural products fluctuates every year. In the absence of standardized price information, large corporations and distribution and purchasing managers (MDs) have no choice but to contact wholesalers and local partners every day to estimate the optimal purchase price and to determine the price by considering various variables. Among agricultural products, the items whose price fluctuates the most are vegetables. Among vegetables, we will predict the distribution price as accurately as possible, focusing on onions, radishes, and Chinese cabbage, which are used as essential ingredients for Korean diet. And using the predicted results will help producers, consumers, distributors, and related organizations, who are consumers, to make decisions about market shipments and sales, such as contract unit price calculation, efficient cultivation method, and harvest time. In addition, price prediction of agricultural products can induce smooth supply and demand control by establishing a sound distribution

order and promote effective price stabilization measures, so that farmers who grow agricultural products can minimize management risks.

In this paper, we aim to find the most appropriate method for predicting price of agricultural products by using deep learning models. The data used in this paper are data for 5 years from 2017 to 2021 of weather data from the Korea Meteorological Agency, consumer price inflation, onion price, radish price, and cabbage price. The collected data was preprocessed according to the deep learning model structure, and then learned and evaluated using RNN, LSTM, and GRU models. Evaluation results Onion RNN model 90.5%, Radishe RNN model 92.9%, Napa cabbage RNN model 91.5%, Onion LSTM model 94.8%, Radishe LSTM model 95.5%, Napa cabbage LSTM model 93.3%, Onion GRU model 93.1, Radishe GRU model 95.5% Napa cabbage GRU model is 92.6%. LSTM Model shows the best performance. This result outperforms the 90% level of accuracy shown by previous time series data prediction and regression analysis models. Through this research, it is possible to establish a sound distribution order for agricultural products, induce supply and demand adjustment, promote effective price safety measures, and increase overall welfare from the standpoint of consumers and producers.

광주대학교
GWANGJU UNIVERSITY

제1장 서론

제1절 연구 배경

국내에서 판매되고 있는 농산물은 정확하게 제공되는 가격이 없어 매년 폭락하거나 급등하는 상황이 자주 일어난다. 유통 중간에 농산물을 구매하는 기업도 정확한 정보가 없어 수십억 원에서 수조 원에 달하는 농산물 구매를 임의로 결정해야 한다. 이에 대기업이나 유통 구매 책임자(MD)는 표준화된 가격 정보 부재로 일일이 도매시장이나 산지 거래처에 연락하고 여러 변수를 고려하여 최적의 구매 가격을 추정하여 가격을 결정하여 진행하기에 가격예측이 어려운 실정이다[1]

농산물 중 가격 변동이 가장 많이 일어나는 품목은 채소류이며, 이중 국민 식생활의 필수 식재료로 활용되고 있는 양파, 무, 배추를 중심으로 유통가격을 정확하게 예측하여 수요자인 생산자, 소비자, 유통업자, 관계기관 등에 신속히 제공함으로써 계약 단가 산정, 효율적인 재배 방식, 수확 시기 등 시장 출하 및 매매에 관한 의사결정에 도움을 주고자 한다. 또한, 농산물 가격예측은 건전한 유통질서를 확립하여 원활한 수급조절을 유도하는 동시에 실효성 있는 가격 안정 대책추진을 할 수 있어서 농산물을 재배하는 농가는 경영 위험을 최소화할 수 있기에 농산물 가격예측은 필요하다.

현재 국내 농산물 가격예측 서비스는 경남 도청에서 주요 농산물 가격예측 시스템을 통해 서비스가 되고 있다. 주요 농산물 가격예측 시스템은 수급(유통)분석 카테고리에서 가격 동향, 수급 상황, 빅데이터를 제공하고 있고, 이중 가격 동향은 주요 농산물인 고추, 깻잎, 마늘, 양파, 양상추, 시금치 등 11개 품목을 제공하고 있다. 농업기상 탭에서는 시군별 기상정보, 주산지 기상정보를 제공하고 있다. 이러한 정보는 경남 도청에서 제공하여 경상남도의 가격만을 예측하고, 기상정보도 경상남도의 시군별 기상정보만 제공하고 있다. 이에 다른 지역의 농산물 가격 정보는 서비스되고 있지 않아 다른 지역의 농가에서는 가격을 예측하여 대응하기 어려운 면이 있다.

본 연구에서는 광주 지역을 대상으로 국민 식생활의 필수 식재료로 활용되고 있

는 채소인 양파, 무, 배추에 대한 가격예측 모델을 연구하고자 한다. 가격예측을 위해 사용된 데이터는 농산물 과거 가격 데이터, 광주 날씨 데이터, 소비자 물가 상승률을 수집하여 분석하였다. 분석 방법은 RNN(Recurrent Neural Network) 계열의 딥러닝(Deep Learning)을 활용하여 농산물 가격을 예측하고, 가격예측 모델의 성능을 비교 분석함으로써 농산물 가격의 예측 성능을 높이고 가장 성능이 좋은 모델을 개발하는 것을 목표로 한다.

제2절 연구 방법

농산물 가격예측 모델을 만들기 위해서는 농산물 가격 변동에 영향을 주는 데이터 수집이 필요하다. 본 논문에서는 농산물 가격예측을 위해 국민 식생활의 필수 식재료로 활용되고 있는 채소류 중 양파, 무, 배추를 선정했다. 가격을 예측하기 위해 수집한 데이터는 가격에 가장 영향을 많이 미치는 농산물의 과거 가격 데이터와 물가 상승률 데이터를 수집하고, 농산물 재배에 가장 큰 영향을 미치는 기상 데이터를 수집한다. 데이터 수집은 농산물 유통정보서비스(KAMIS), 기상청 날씨 데이터 서비스 기상자료개방포털, 국가통계 포털(KOSIS)에서 진행했다. 데이터 세부 항목으로는 양파 가격, 무 가격, 배추 가격, 기온, 강수량, 습도, 풍량, 소비자 물가 상승률 데이터이며, 각 항목은 2017년부터 2021년까지 5년간의 데이터를 수집하였다.

이렇게 수집한 데이터는 시간 순서로 나열된 특징을 가진 데이터로 가격예측 시 과거의 정보를 반영하여 미래를 예측할 수 있는 딥러닝 알고리즘을 사용해야 한다. 순차적인 의미 정보를 처리하여 과거의 정보를 반영하고, 미래를 예측할 수 있는 딥러닝 모델은 대표적으로 RNN모델이 있으며, 최근 RNN의 단점을 보완하여 발전된 모델로는 LSTM, GRU 등이 있다. RNN 계열의 딥러닝 모델에 데이터의 시간 정보를 반영할 수 있도록 전처리하여 가공한 후 훈련데이터 세트(Training Data Set)과 테스트데이터 세트(Test Data Set)로 나누어 딥러닝 모델을 학습한다. 모델 학습은 앞서 말한 RNN, LSTM, GRU 모델을 사용하여 학습한 후 모델 성능을 비교 분석하였다.

본 논문에서는 Python언어를 기반으로 오픈소스 라이브러리인 사이킷런(skikit-learn) 1.0버전, 텐서플로우(TensorFlow) 2.6버전과 텐서플로우 내부에 있는 케라스(keras)를 사용하였다. 최종 모델 평가는 손실함수 RMSE(Root Mean Squared Error)와 백분율 오차율(%), 정확도(%)를 사용하여 각 모델 성능을 비교하였다. 농산물 가격예측 모델 구현을 위한 학습데이터로는 2017년 1월 2일 부터 2020년 12월 31일까지 4년간의 데이터를 사용했고, 구현된 모델의 성능을 검증하기 위해서는 2021년 1월 1일 부터 2021년 12월 30일까지 1년간의 데이터로 가격을 예측한 후 실제 농산물 가격과 비교하였다.



제2장 관련 연구

제1절 기존 농산물 가격예측

본 절에서는 농산물 가격예측 연구를 위해 기존에 연구되었던 농산물 가격예측 방법을 확인함으로써 연구 동향을 파악하고, 연구 방법을 분석하여 기존 연구를 발전시킬 방안을 모색한다.

2021년에 연구된 ‘Lasso 회귀분석을 활용한 농산물 가격예측 모델 변수 선정 연구’는 채소류의 양파와 대파의 소매 가격을 예측하였으며, 예측하기 위한 데이터로는 전국 날씨, 지수, 수출량, 수입량, 수출금액, 수입금액, 재배면적, 생산량, 생산 시기를 사용하였다[2]. 이 연구는 선형회귀모델로 현실적인 데이터 분석 시 낮은 정확도를 개선하기 위해 별점함수를 추가하여 분석하는 Lasso 모델을 사용하여 예측하는 방법으로 예측하였다. 연구 결과는 선행연구 대비 양파의 RMSE는 0.22만큼 성능이 낮아지고, 대파는 0.07만큼 성능이 우수했다. 이 연구는 대파의 성능은 향상되었지만, 양파의 성능은 오히려 낮아진 결과가 도출되어 다양한 농산물의 성능도 향상할 수 있는 연구가 필요하다.

[표 2-1] Lasso 회귀분석을 활용한 농산물 가격예측 모델 연구 결과

RMSE	예측기간	1일 후	3일 후	5일 후	7일 후
양파	기본 변수	0.2457	0.2277	0.2326	0.2775
	Lasso 변수	0.2270	0.2004	0.2239	0.2406
대파	기본 변수	0.0712	0.0792	0.0747	0.0664
	Lasso 변수	0.0706	0.0676	0.0765	0.0691

2018년에 연구된 ‘LSTM 네트워크를 활용한 농산물 가격예측 모델’은 대과, 양과, 애호박, 쌀, 시금치 5가지 농산물을 대상으로 소매 가격을 예측하고, 태풍과 같은 자연재해가 농산물 가격에 미치는 영향을 예측한다[3]. 이 연구의 전국 지역을 대상으로 가격을 예측하였으며, 사용한 데이터는 기온, 강수량, 습도, 풍량, 적설량, 유가, 물가 상승률, 전년 수확량 등 108개의 데이터를 사용하였다. 가격예측에 사용된 딥러닝 알고리즘은 의미 정보를 처리하여 과거의 정보를 반영할 수 있는 LSTM을 사용하였다. LSTM을 사용하기 위해 데이터는 21일을 기준으로 학습데이터 포맷을 만들어 108 x 21 형태로 LSTM 모델 학습에 사용된다. 연구 결과 모델의 정확도는 백분율 기준으로 93%가 나와 기존 예측모델이 보여주는 90% 수준보다 높은 정확도를 보인다. 해당 논문은 정확도는 높게 나왔으나 가격을 예측하기 위해서는 예측하고자 하는 날짜를 기준으로 108개의 데이터가 필요하여 현실적으로 해당 모델을 사용하여 가격을 예측하고 사용하기에는 어려움이 보인다.

[표 2-2] LSTM 네트워크를 활용한 농산물 가격예측 모델 정확도

지역	대과	양과	애호박	쌀	시금치	전체 합계
대구	0.047002	0.059741	0.098147	0.056684	0.085620	0.069439
광주	0.053291	0.050698	0.090949	0.059716	0.071724	0.065276
서울	0.054660	0.072942	0.077357	0.052059	0.070472	0.065498
대전	0.057111	0.053470	0.121505	0.059265	0.074336	0.073137
부산	0.051233	0.082001	0.096845	0.060045	0.076212	0.073267
합계	0.052659	0.06377	0.096961	0.057554	0.075673	0.069323

2016년에 연구된 ‘인공신경망 기법을 이용한 최적의 농산물 가격예측모델 개발’은 인공신경망 은닉층 구성 시 은닉층의 수, 모멘텀과 학습률 계수를 조절하면서 가격예측 정확도를 비교하였다[4]. 이 연구는 가격예측을 위해 수집된 21개의 데이터 중 연관성이 높다고 판단된 7개의 데이터를 선정하고, 단순 인공신경망 기법을 사용하여 분석하였다. 인공신경망은 은닉층 수 17개, 모멘텀 값 0.5, 학습률 0.4일 때 예측정확도가 가장 높았으며, 선행연구보다 14.77% 상승했다. 이 논문은 단순 인공신경망만을 사용하여 최근 많이 사용되고 있는 RNN 모델 기반의 딥러닝 모델을 사용하면 더 좋은 성능을 보일 것으로 보인다.

[표 2-3] 인공신경망 기법을 이용한 최적의 농산물 가격예측모델 정확도

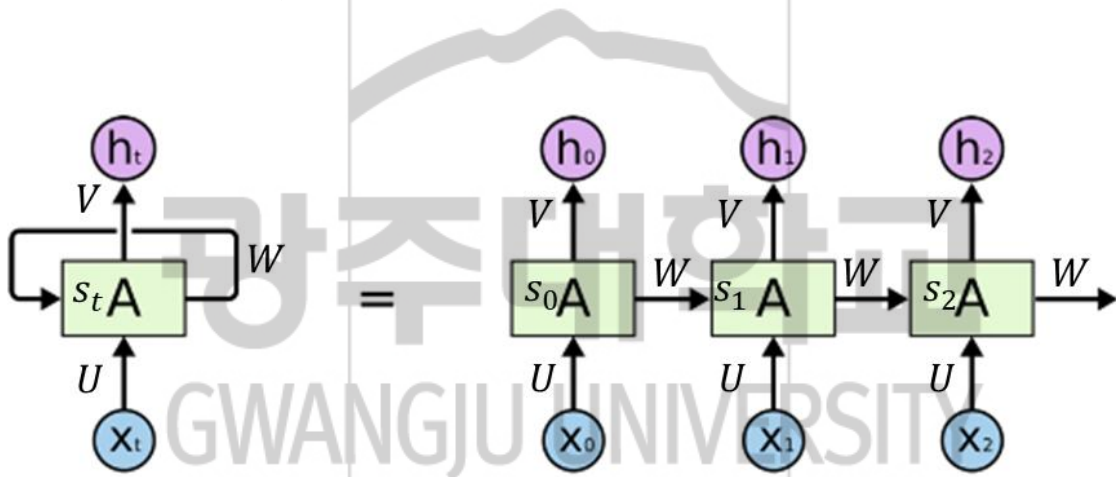
구분	평균 가격	최소 가격	최대 가격
인공신경망 모델을 통한 가격	27168.92	12196.81	54152.71
실제 가격(원)	20346.81	9333	43333
정확도(%)	74.89	76.52	80.02

기존 농산물 가격예측 연구를 분석한 결과 농산물은 채소류가 가장 많이 분석되었으며, 농산물 가격예측을 위해 기온, 강수량, 습도, 풍량, 적설량, 계절 변화, 수요량, 공급량, 과거 가격, 유가, 물가, 전국 생산량, 일일 가격 표준편차, 재배면적, 농가소득 등의 데이터가 사용된다는 것을 알 수 있었다. 또한, 지역에 대한 특성이 반영된 연구가 없어서 전국 농산물 가격은 예측할 수 있으나 지역별 가격예측은 불가능하여 실제 활용되기 어려울 것으로 보인다. 이에 본 논문은 광주광역시로 한정 지어 채소류인 양파, 무, 배추에 대한 가격을 예측하고자 한다. 사용된 데이터는 양파, 무, 배추 가격, 광주광역시 지역에 대한 기온, 상대습도, 풍속, 소비자 물가 상승률 데이터를 활용하여 딥러닝 알고리즘을 연구하여 기존 선행 연구된 성능보다 높은 성능을 보이는 모델을 연구한다.

제2절 농산물 가격예측을 위한 딥러닝 기법

1. RNN(Recurrent Neural Network) Model - 순환신경망 모델

기존 인공신경망(ANN; Artificial Neural Network)은 문장의 문맥을 파악하기 위해 문장의 흐름을 이해할 수 있는 기능을 구현하기에는 어려움이 있다. 데이터의 흐름을 이해할 수 없는 인공신경망(ANN)은 시간의 흐름을 이해해야 하는 자연어 처리, 시계열 데이터 처리 분야에서 정확도가 낮은 편이다. 과거의 정보를 가지고 미래를 예측할 수 있도록 시간 개념을 더해 개발된 딥러닝 알고리즘이 RNN(Recurrent Neural Network) 모델이며[5], RNN 모델은 루프를 통해 순환할 수 있는 구조로 구성되어 있다. RNN 모델은 [그림 2-1]처럼 하나의 네트워크가 여러 개 복사된 형태를 띠고 있으며, 각각의 네트워크는 자신이 가진 정보를 다음 네트워크로 넘겨 준다.



[그림 2-1] RNN 구조

[그림2-1]에서 x_t 는 t 시점의 입력값이며, s_t 는 t 시점의 히든 스테이트(Hidden State)를 나타낸다. 히든 스테이트인 s_t 의 계산 식은 (1)과 같다.

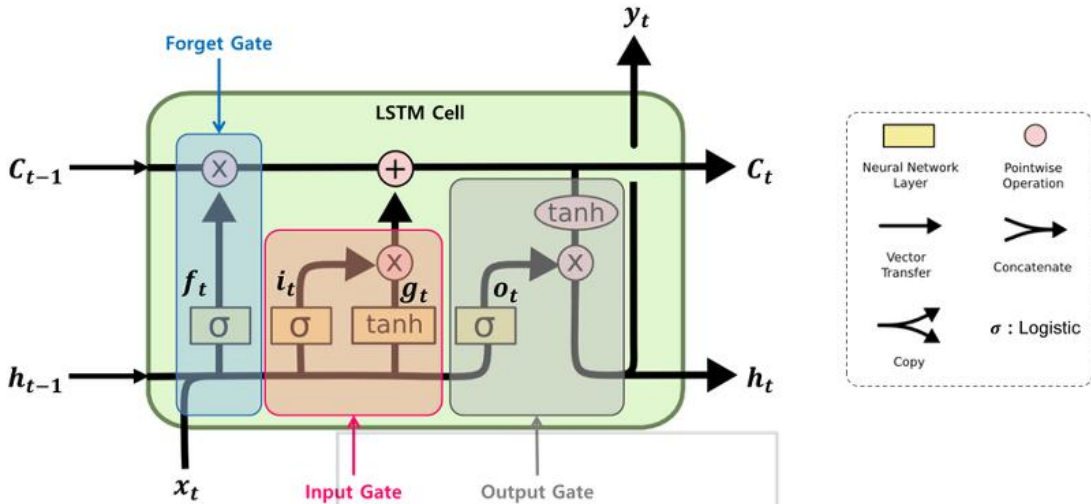
$$s_t = f(Ux_t + Ws_{t-1}) \quad (1)$$

s_t 는 이전 단계($t-1$)의 히든 스테이트와 현재 입력(t) 단계를 기반으로 계산된다. 함수 f 는 활성화 함수 \tanh , ReLU와 같은 비선형 변환을 의미하며, U, W, V 는 모든 시점에서 공유된다. x_t 는 입력 데이터(input data)를 의미하며 본 논문에서는 가격 데이터, 기상 데이터 등 시계열 구조를 가진 데이터이다. H_t 는 출력 데이터(output data)를 의미하며, 농산물 가격예측 결과이다.

RNN의 히든 스테이트는 과거 정보를 현재, 미래에 전달함으로써 RNN의 핵심 기능이라고 볼 수 있다. 그러나 현실 데이터를 반영할 경우 기울기 소실 문제(Vanishing Gradient Problem)로 학습 성능이 낮다는 단점이 있다. 기울기 소실 문제는 RNN 모델 구조로는 해결할 수 없다는 연구 결과가 있다[8]. 이러한 문제를 해결하기 위해 LSTM, GRU와 같은 다양한 모델이 개발되었다.

2. LSTM(Long Short Term Memory) Model - 장단기 메모리 모델

LSTM(Long Short Term Memory, 장단기 메모리) 모델은 RNN 모델의 기울기 소실 문제를 해결하기 위해 고안된 딥러닝 알고리즘이다.[11] LSTM 모델은 망각 게이트(forget gate)라 부르는 게이트가 핵심이다. 망각 게이트를 통해 역전파(Back Propagation) 계산을 하면, 기울기 값이 사라지거나 급격히 증가하는 문제를 방지하여 기울기 소실 문제를 해결할 수 있다. 이로 인해 LSTM 모델은 먼 과거의 정보까지 반영하여 학습할 수 있게 된다. LSTM 모델 구조는 RNN 모델과 같이 순환 구조로 되어있으나 내부노드 형태가 [그림 2-2]와 같이 상대적으로 복잡하다.



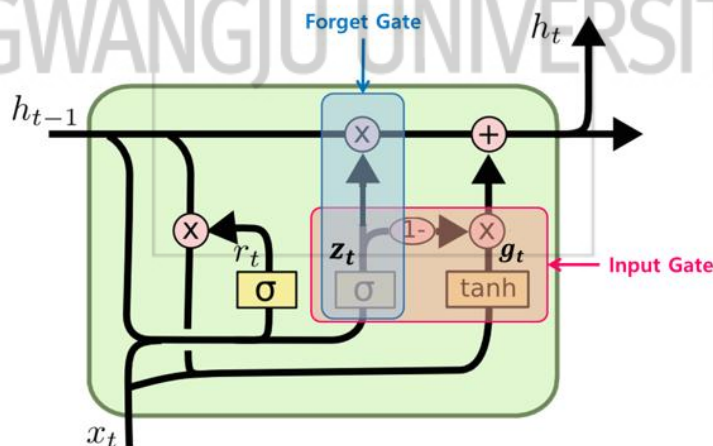
[그림 2-2] LSTM 내부구조

LSTM 모델은 내부에서는 어떤 정보를 버릴지, 어떤 새로운 정보를 저장할지, 어떤 출력값을 출력할지 결정한다. 첫 번째로 어떤 정보를 버릴지 선택하는 과정은 망각 게이트 레이어(forget gate layer)라고 부르는 시그모이드 레이어(sigmoid Layer)로 구성되어 있다. 망각 게이트(forget gate)를 통과한 출력 값은 h_{t-1} 과 x_t 를 입력 값으로 받으며, 0과 1 사이의 값으로 구성되고 0에 가까울수록 값을 버리고, 1에 가까울수록 그대로 유지하라는 뜻을 가진다.[12] 두 번째 새로운 정보를 저장할지 확인하는 과정은 input gate layer라고 부르는 시그모이드 레이어는 어떤 값을 업데이트할지 결정하고, 하이퍼볼릭탄젠트 레이어(tanh Layer)는 셀 스테이트에 더해질 수 있는 새로운 후보 값을 만들어 낸다. 이 두 가지 값을 합쳐서 다음 스테이트에 영향을 준다. 마지막 세 번째는 어떤 출력값을 출력할지 결정하는 과정은 이전 히든 스테이트(hidden state) 값과 현재 입력 값을 이용하여 출력 게이트(output gate) 값을 계산하고, 출력 게이트 값과 현재 입력 값을 결합하여 현재의 히든 스테이트 값을 계산한다. LSTM 모델의 단계별 수식은 식 (2)와 같다[6, 9].

$$\begin{aligned}
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
\tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
h_t &= o_t * \tanh(C_t)
\end{aligned} \tag{2}$$

3. GRU(Gated Recurrent Units) Model - 게이트 순환 유닛 모델

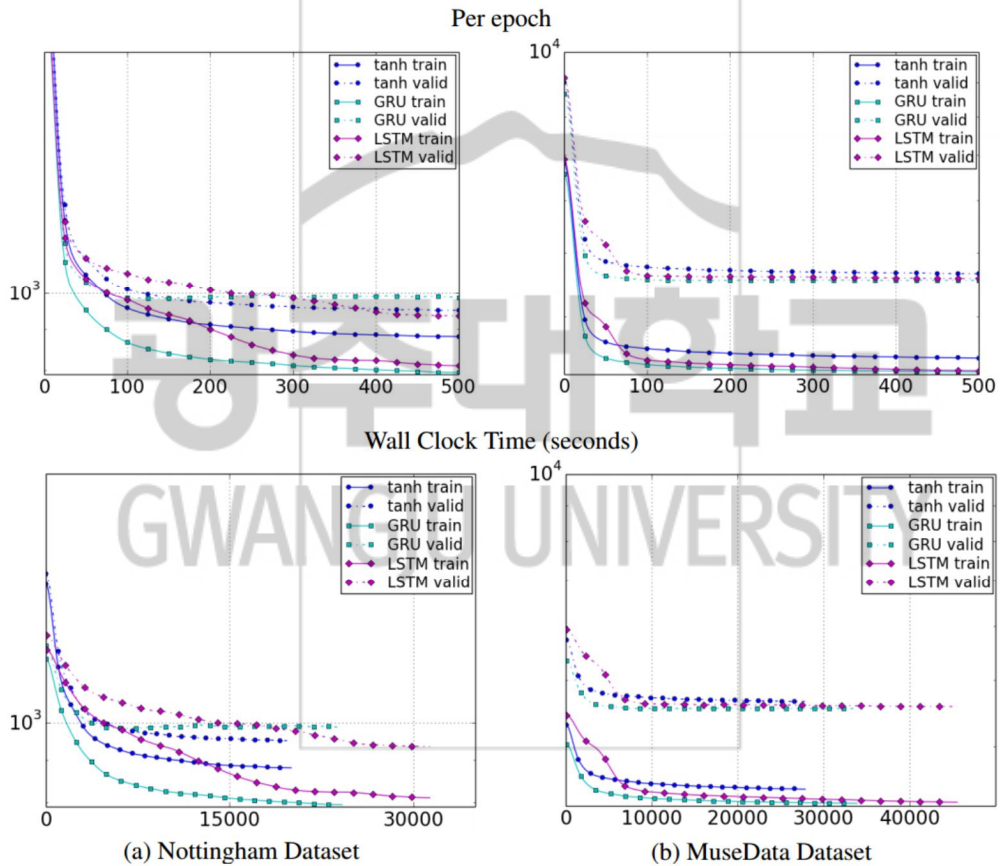
GRU(Gated Recurrent Units, 게이트 순환 유닛) 모델은 복잡했던 LSTM 모델보다 계산량을 줄여 개발되었다. GRU 모델은 리셋 게이트(reset gate)와 업데이트 게이트(update gate) 두 개의 게이트로 구성되며, LSTM 모델처럼 메모리를 보호한다. 기존 LSTM 모델의 망각 게이트와 입력 게이트를 통합하여 하나의 업데이트 게이트로 구성하고, 셀 스테이트와 히든 스테이트를 하나로 통합하였다. 이로 인해 GRU 모델은 LSTM 모델에 비해 파라미터수가 적고, 연산 비용이 적게 들고, 구조도 간단해진다. 다만 성능 면에서만 보면 LSTM 모델과 비슷하거나 떨어진다[8].



[그림 2-3] GRU 내부구조

$$\begin{aligned}
 r_t &= \sigma(W^{(r)}x_t + U^{(r)}h_{t-1}) \\
 z_t &= \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \\
 \tilde{h}_t &= \tanh(Wx_t + r_t \cdot Uh_{t-1}) \\
 h_t &= z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t
 \end{aligned} \tag{3}$$

RNN 모델보다 LSTM, GRU 모델의 성능이 우수하지만, LSTM과 GRU 중에서는 어떤 모델이 더 좋은지 결론이 나지 않은 상태라서 연구원들은 학습하는 컴퓨터 성능을 고려하여 사용한다[9].



[그림 2-4] 반복횟수, 계산시간을 고려한 학습 곡선

제3장 딥러닝을 활용한 농산물 가격예측 모델 연구

제1절 실험 데이터

1. 데이터 수집

본 연구에서는 농산물 가격예측을 위해 기상청 날씨 데이터와 소비자 물가 지수 데이터를 사용하였다. 기상청 날씨 데이터는 기상청 날씨 데이터 서비스 기상자료 개방포털에서 2017년부터 2021년까지 총 5년간의 광주 기상정보데이터를 사용하였다. 데이터는 csv(comma-separated values)형태의 파일이며, 일시, 지점명, 평균기온(°C), 일 강수량(mm), 평균 풍속(m/s), 평균 상대습도(%)의 6개의 속성과 1,825개의 데이터로 이루어져 있다. 이 속성 중 일 강수량(mm)은 60% 이상 결측치이기 때문에 농산물 가격예측에는 사용하지 않고 삭제처리 하였다.

[표 3-1] 기상청 날씨 데이터

일시	지점명	평균기온 (°C)	일 강수량 (mm)	평균풍속 (m/s)	평균 상대습도 (%)
2017-01-01	광주	4.5		0.7	80.6
2017-01-02	광주	7.2	0	1	81.6
2017-01-03	광주	5.7		1	68.3
2017-01-04	광주	5		0.8	77.1
2017-01-05	광주	6.8	0	1.6	71.3
...					
2021-12-27	광주	-2.1		1.6	65.9
2021-12-28	광주	3.3		0.6	69.3
2021-12-29	광주	4.5		1.3	70.4
2021-12-30	광주	1.9	1	2	67.1
2021-12-31	광주	-1.9	0	2.1	62.5

소비자 물가 지수 데이터는 국가통계 포털(KOSIS)에서 2017년부터 2021년까지 총 5년간의 전국 소비자 물가 데이터를 사용하였다. 데이터는 csv형태의 파일이며, 소비자 물가 지수는 월 단위로 구성되어 60개의 데이터로 이루어져 있다.

[표 3-2] 소비자 물가 데이터

일시	전국		일시	전국
2017.01	97.366	...	2021.06	102.05
2017.02	97.632		2021.07	102.26
2017.03	97.565		2021.08	102.75
2017.04	97.442		2021.09	103.17
2017.05	97.546		2021.10	103.35
2017.06	97.338		2021.11	103.87
2017.07	97.499		2021.12	104.04

학습에 필요한 데이터 중 목표값(label)인 농산물 가격 데이터는 농산물 유통정보서비스(KAMIS)에서 제공한 채소류의 양파, 무, 배추 가격을 2017년부터 2021년까지 광주 지역의 데이터를 사용하였다. 가격예측 성능을 측정하기 위해 2017년부터 2020년까지 4년간의 데이터와 2021년 데이터를 나누고, 4년간의 데이터로 딥러닝 모델을 학습하여 2021년 가격을 예측한다.

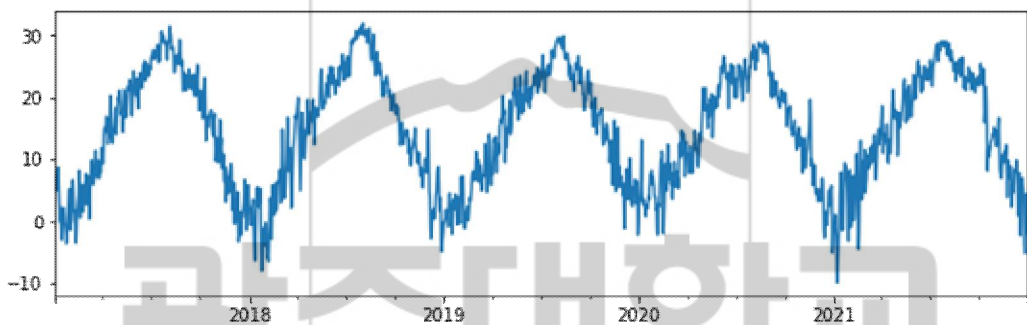
[표 3-3] 양파, 무, 배추 가격 데이터

일시	양파	무	배추
2017.01.02	15,000	22,000	8,000
2017.01.03	15,000	23,000	8,000
2017.01.04	15,000	23,000	8,000
2017.01.05	16,500	24,000	8,000
2017.01.06	16,500	24,000	8,000
...			
2021.12.27	13,650	11,200	9,120
2021.12.28	13,650	11,200	8,120
2021.12.29	13,650	11,200	9,120
2021.12.30	13,650	11,200	9,120
2021.12.31	13,650	11,200	9,120

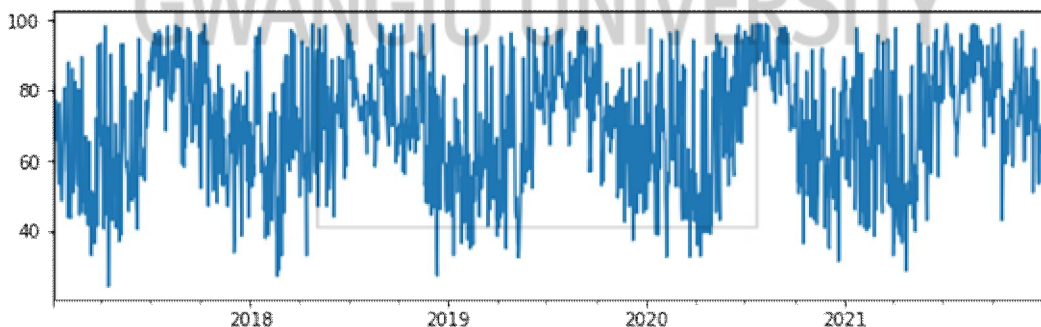
2. 데이터 분석

기상청 날씨 데이터와 소비자 물가 지수 데이터, 농산물 가격 데이터는 총 9개 속성(지점명, 평균기온, 일 강수량, 평균 풍속, 평균 상대습도, 물가, 양파 가격, 무 가격, 배추 가격)으로 이루어져 있다. 이 중 기상청 날씨 데이터의 지점명은 광주

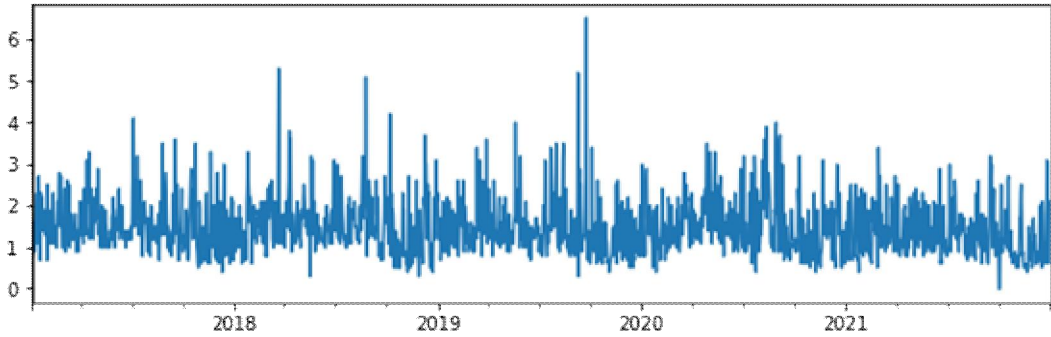
지역 하나이므로 삭제하고, 일 강수량은 60%이상 결측치를 가지므로 삭제하였다. 다음은 시간의 흐름과 농산물 가격 변동의 관계를 알기 위해서 기상청 날씨 데이터와 소비자 물가 지수 데이터, 농산물(양파, 무, 배추)의 가격 데이터를 시각화하였다. [그림 3-1], [그림 3-2], [그림 3-3]은 2017년부터 2021년까지의 평균기온, 평균 상대습도, 평균 풍속 그래프이다. 평균기온과 평균 상대습도는 1년을 주기로 반복되는 것을 볼 수 있다. 기온의 경우 5년간 최저 기온은 -9.9°C , 최고 기온은 32°C 이며, 21년 겨울철(12월, 1월, 2월)에는 기온이 $2^{\circ}\text{C}\sim 5^{\circ}\text{C}$ 가량 더 떨어진 것을 확인할 수 있다. 상대습도의 경우 최저 상대습도 23.9%, 최고 상대습도 99%이며 상대적으로 2021년 6월 이후 평균 상대습도가 증가 되었다. 풍속의 경우 특별한 주기를 가지고 있지 않으나 0m/s에서 3m/s까지 반복되고 있다. [그림 3-4]은 풍속의 변화를 자세히 확인하기 위해 60일간의 평균 풍속 변화를 보여준다.



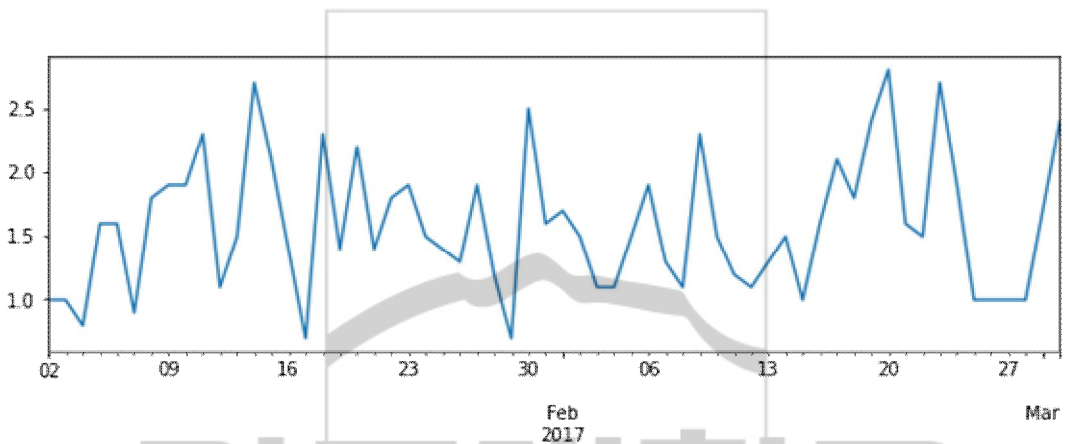
[그림 3-1] 기상청 날씨 데이터 - 평균기온



[그림 3-2] 기상청 날씨 데이터 - 평균 상대습도

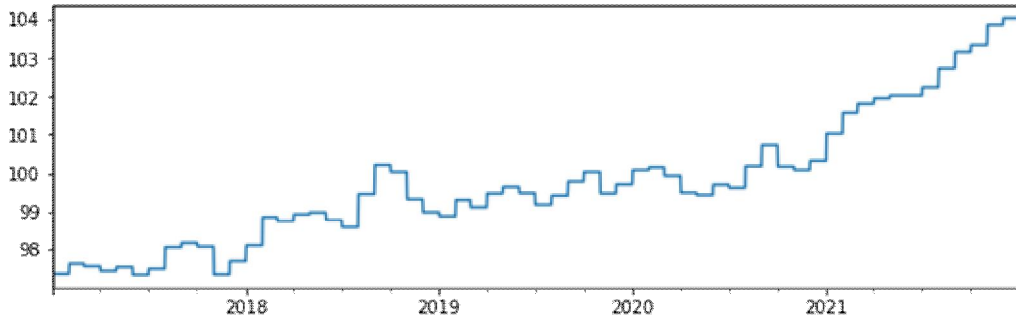


[그림 3-3] 기상청 날씨 데이터 - 평균 풍속



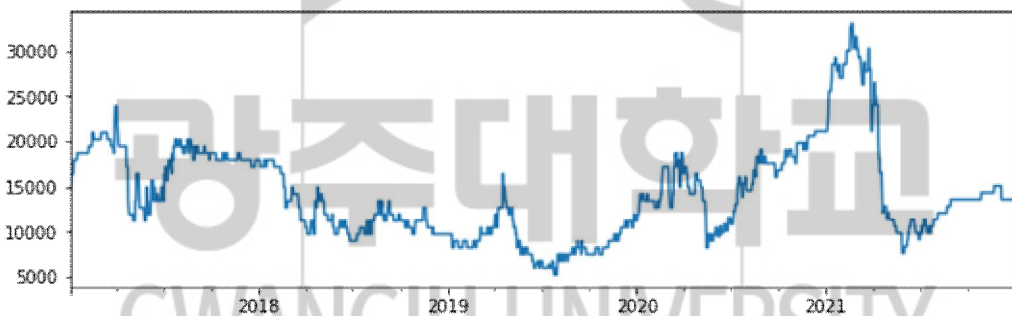
[그림 3-4] 기상청 날씨 데이터 - 60일간 평균 풍속

[그림 3-5]는 2017년부터 2021년까지 소비자 물가 데이터 그래프다. 소비자 물가는 증가와 감소를 반복하지만, 해를 거듭할수록 물가가 상승하는 모습을 보인다. 특히 최근 2021년에는 급격하게 물가가 상승하고 있다. 2021년에 물가가 상승한 이유는 코로나19 확산으로 위축된 경기를 살리기 위해 개인과 기업에 지원금을 지급하거나 대출을 지원하는 등 많은 돈이 풀려 시중에 사용할 수 있는 돈, 통화량이 증가했다. 또한, 코로나19 백신 접종으로 경기가 회복될 것이라는 기대감이 높아지면서 경제 전반에 걸쳐 수요가 빠른 속도로 늘어나게 되고, 이에 맞춰 기업은 개인 소비에 맞춰 생산을 늘리고, 생산을 늘리기 위해서는 많은 원자재가 필요하다. 이렇게 늘어나는 수요를 공급이 따라가지 못해 원자재 가격이 급등하여 전체적인 소비자 물가가 상승한 것으로 보인다.



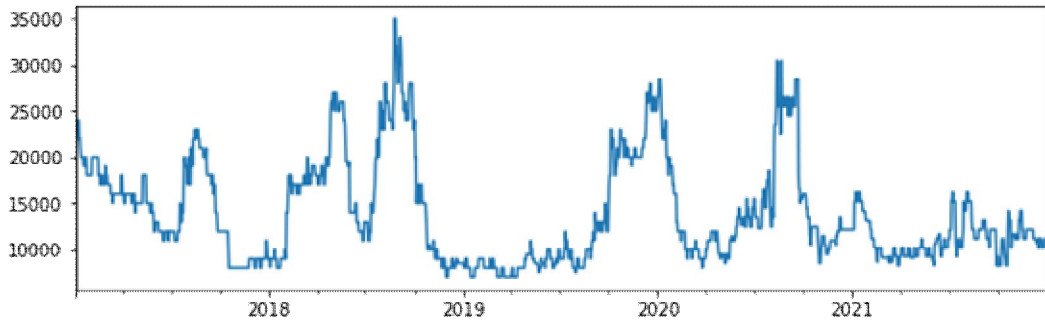
[그림 3-5] 소비자 물가 데이터

[그림 3-6]은 2017년부터 2021년까지 광주 지역의 양파 가격 데이터 그래프다. 양파 가격은 2017년 15,000원을 시작으로 2019년 6월까지 하강 곡선을 보이고, 최저가는 5,250원이다. 그 이후 2021년 코로나19로 인해 급격히 증가하여 최고가 33,150원을 찍은 후 약 13,000원대로 안정화된 모습을 보인다. 2021년 급격한 변화를 보아 소비자 물가 상승이 영향을 미친 것으로 보인다.



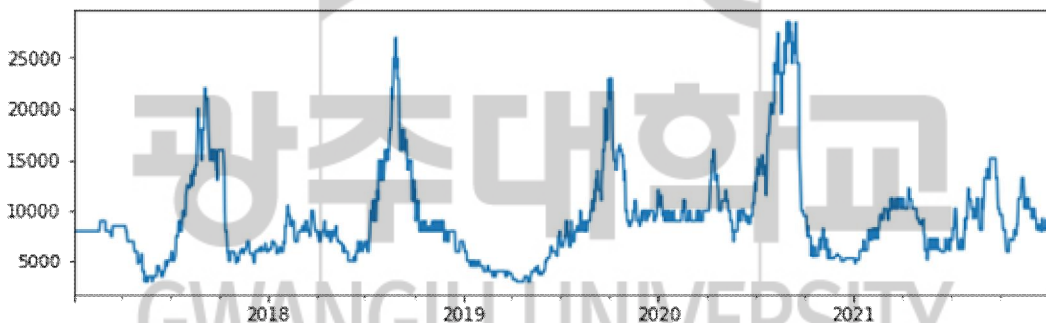
[그림 3-6] 농산물 양파 가격 데이터

[그림 3-7]은 2017년부터 2021년까지 광주 지역의 무 가격 데이터 그래프다. 무 가격은 2017년 22,000원으로 시작하고, 2018년에는 7,000원대까지 떨어지는 모습을 보인다. 무 가격은 평균 14,000원을 기록하고 있으며, 5년간 최고가는 35,000원이며 최저가와 5배 차이가 나는 것을 확인할 수 있다. 특히 4분기(10월, 11월, 12월)에 급격한 가격 변동이 일어난다. 4분기에 가격 변동이 큰 이유는 11월 하순부터 12월 상순까지 김장철로 무가 가장 많이 사용되는 시기이다.



[그림 3-7] 농산물 무 가격 데이터

[그림 3-8]은 2017년부터 2021년까지 광주 지역의 배추 가격 데이터 그래프다. 배추 가격은 2017년 8,000원으로 시작하고, 평균 9,210원이다. 4분기(10월, 11월, 12월)에 급격한 가격 변동이 일어나는 것을 확인할 수 있다. 급격한 가격 변동이 일어나는 이유는 무와 마찬가지로 11월 하순부터 12월 상순까지 김장철로 인해 배추가 많이 사용되는 시기이기 때문이다.



[그림 3-8] 농산물 배추 가격 데이터

이처럼 소비자 물가, 날씨 데이터, 날짜(김장철 시기)가 양파, 무, 배추 가격에 많은 영향을 미치는 것을 확인할 수 있다. 본 논문에서는 이러한 내용을 잘 반영하기 위해 시간 순서로 데이터를 전처리하고, 시간 정보를 반영할 수 있는 RNN, LSTM, GRU 딥러닝 모델을 사용하여 분석한다.

3. 데이터 전처리

본 논문의 데이터 전처리는 데이터 분석한 내용을 기반으로 수집한 데이터 중 학습에 사용할 속성을 선택하고, 날짜별로 데이터를 병합한 후 시간 순서를 반영할 수 있도록 데이터의 구조를 변경한다.

기상청 날씨 데이터는 일시, 지점명, 평균기온(°C), 일 강수량(mm), 평균 풍속(m/s), 평균 상대습도(%) 데이터 속성을 가지고 있으며, 이 중 일 강수량(mm)은 60%이상 NaN 데이터로 구성되어 있다. NaN은 데이터가 존재하지 않은 경우를 말한다. 일 강수량 데이터를 자세히 분석하면 NaN 데이터는 비가 오지 않은 날인 것을 볼 수 있다. 이런 경우 NaN은 0으로 값을 대체하여 분석할 수 있다. 그러나 데이터를 자세히 살펴보면 NaN이 아닌 다른 데이터에 0으로 표시된 데이터가 많아서 NaN과 0의 차이점이 있을 것으로 보인다. 또한, NaN은 데이터 수집 시 오류가 발생하여 누락된 기록일 수도 있다. 이에 일 강수량 속성은 오류를 최소화하기 위해 삭제하였다.

일시	평균기온(°C)	평균 풍속(m/s)	평균 상대습도(%)
2017-01-02	7.2	1.0	81.6
2017-01-03	5.7	1.0	68.3
2017-01-04	5.0	0.8	77.1
2017-01-05	6.8	1.6	71.3
2017-01-06	8.2	1.6	65.9
...
2021-12-27	-2.1	1.6	65.9
2021-12-28	3.3	0.6	69.3
2021-12-29	4.5	1.3	70.4
2021-12-30	1.9	2.0	67.1
2021-12-31	-1.9	2.1	62.5

[그림 3-9] 기상청 날씨 데이터

국가통계 포털(KOSIS)에서 수집한 소비자 물가 데이터는 월별로 구성되어 있다. 월별로 나와 있는 소비자 물가를 데이터 분석하기 위해서는 기상청 날씨 데이터와 같이 일별로 구조를 변경한 후 병합해야 한다. 예를 들어 월별로 구성된 2017년 01월 소비자 물가 97.366는 일별 데이터로 변환 시 2017년 01월 02일부터 2017년 01월 31일까지 일별 날짜를 생성한 후 2017년 01월에 해당하는 97.366 데이터를 넣는다. 이와 같이 2017년 1월부터 2021년 12월까지 모든 월에 적용하여 기존 60개로 구성된 데이터는 1,825개 데이터로 증가하였다.

시도별	전국	물가
2017.01	97.366	2017-01-02 97.366
2017.02	97.632	2017-01-03 97.366
2017.03	97.565	2017-01-04 97.366
2017.04	97.442	2017-01-05 97.366
2017.05	97.546	2017-01-06 97.366
2017.06	97.338	2017-01-07 97.366
2017.07	97.499	2017-01-08 97.366
2017.08	98.058	2017-01-09 97.366
2017.09	98.172	2017-01-10 97.366
2017.10	98.077	2017-01-11 97.366

[그림 3-10] (좌)월별 소비자 물가 데이터, (우)일별 소비자 물가 데이터

농산물 유통정보 서비스(KAMIS)에서 수집한 농산물(양파, 무, 배추) 가격 데이터는 날짜가 2017-01-02처럼 구성되어 있지 않고 년도 없이 01-02처럼 월과 일로만 구성이 되어있다. 이 문제를 해결하기 위해 데이터 수집 시 연도별로 수집하고, 연도별 csv파일을 load하여 연도를 붙여주었다. 또한, 가격 데이터는 중간중간 날짜가 비어있는데 비어있는 날짜는 주말과 공휴일에 해당한다. 이를 해결하기 위해 비어있는 날짜를 새롭게 생성하고, 가격은 앞날과 뒷날 가격의 평균을 구해 채워 넣어주었다. [그림 3-11]은 전처리가 완료된 농산물 가격 데이터, 기상청 날씨 데이터, 소비자 물가 데이터를 병합한 결과이다.

	양파	무	배추	평균기온(°C)	평균 풍속(m/s)	평균 상대습도(%)	물가
2017-01-02	15000.0	22000.0	8000.0	7.2	1.0	81.6	97.366
2017-01-03	15000.0	23000.0	8000.0	5.7	1.0	68.3	97.366
2017-01-04	15000.0	23000.0	8000.0	5.0	0.8	77.1	97.366
2017-01-05	16500.0	24000.0	8000.0	6.8	1.6	71.3	97.366
2017-01-06	16500.0	24000.0	8000.0	8.2	1.6	65.9	97.366

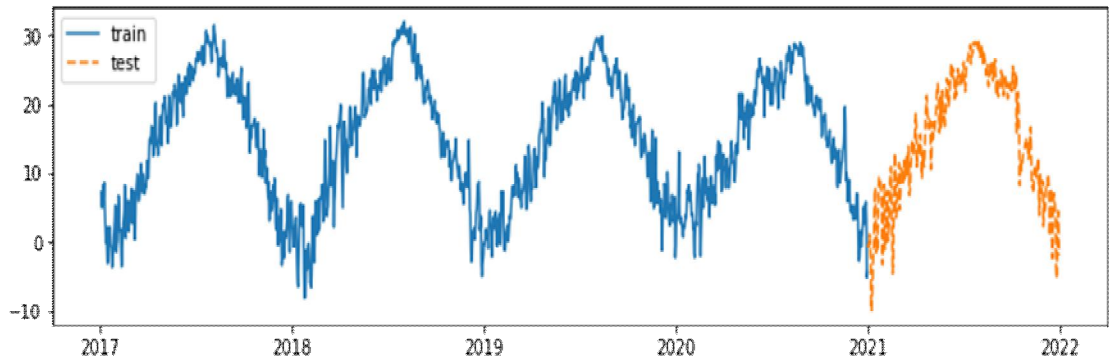
[그림 3-11] 농산물 가격예측을 위한 데이터 셋

양파, 무, 배추 가격을 각각 예측하기 위해 병합한 데이터 셋을 그대로 사용하지 않고 양파, 무, 배추 이렇게 3가지로 나눠서 분석한다. 양파 가격예측을 위해 데이터 속성은 양파, 평균기온, 평균 풍속, 평균 상대습도, 물가로 [그림 3-12]처럼 구성한다. 무와 배추도 마찬가지로 데이터 속성은 농산물(무, 배추), 평균기온, 평균 풍속, 평균 상대습도, 물가로 구성하여 총 3개의 데이터 셋을 구성한다.

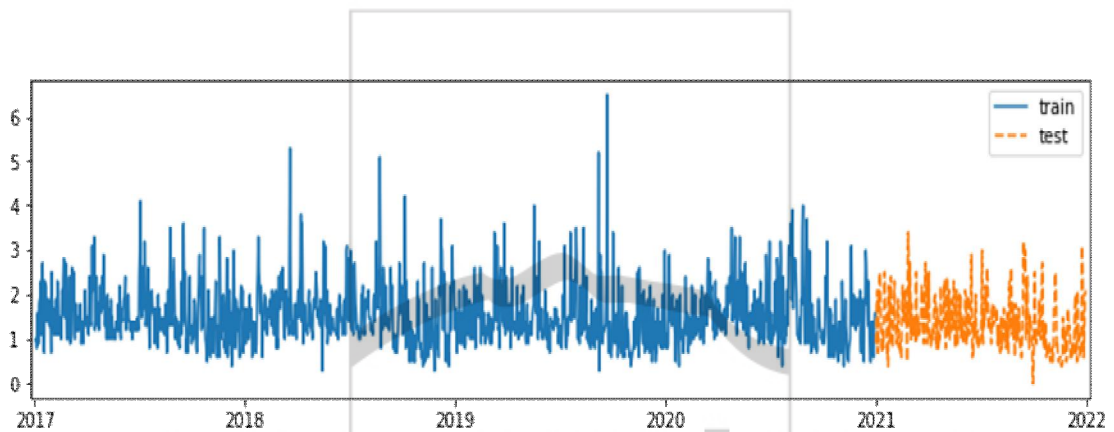
	양파	평균기온(°C)	평균 풍속(m/s)	평균 상대습도(%)	물가
2017-01-02	15000.0	7.2	1.0	81.6	97.366
2017-01-03	15000.0	5.7	1.0	68.3	97.366
2017-01-04	15000.0	5.0	0.8	77.1	97.366
2017-01-05	16500.0	6.8	1.6	71.3	97.366
2017-01-06	16500.0	8.2	1.6	65.9	97.366

[그림 3-12] 양파 가격예측을 위한 데이터 셋

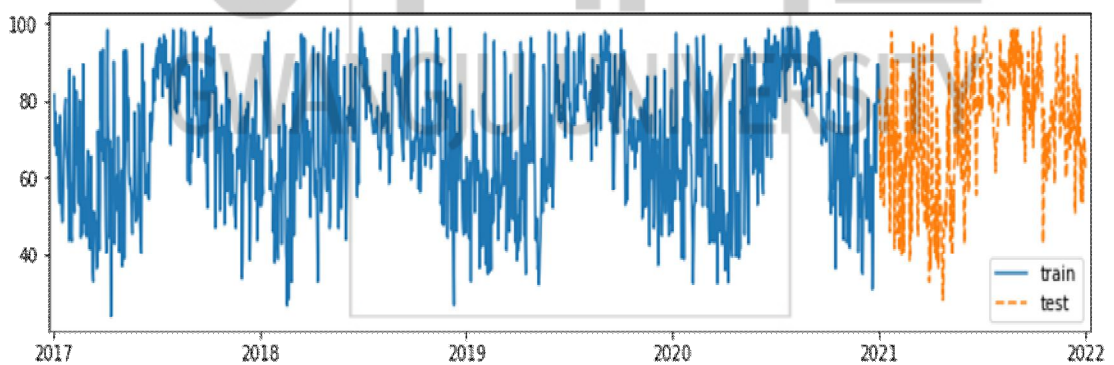
3개의 데이터 셋은 딥러닝 모델을 학습하고 가격예측을 하기 위해 훈련 집합(Training Data Set)과 테스트 집합(Test Data Set)으로 나눈다. 농산물 가격예측은 과거의 정보(기상, 물가, 가격)를 보고 미래의 가격을 예측하는 것을 목적으로 데이터는 시간 순서로 나눠야 한다. Training Data Set과 Test Data Set은 2020년을 기준으로 나눠 2017년 01월 02일부터 2020년 12월 31일까지 Training Data Set으로 구성하고, 2021년 01월 01일부터는 Test Data Set으로 구성한다. [그림 3-13] ~ [그림 3-19]는 Training Data Set과 Test Data Set 분할결과 그래프이다.



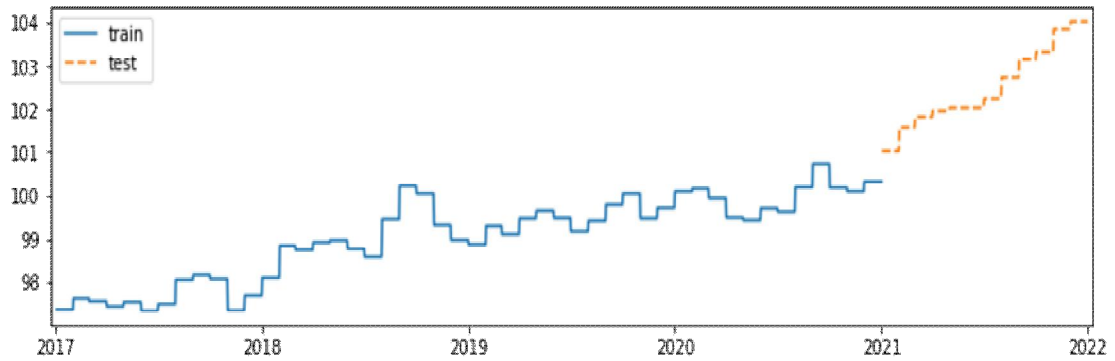
[그림 3-13] 평균기온(°C) train, test 데이터



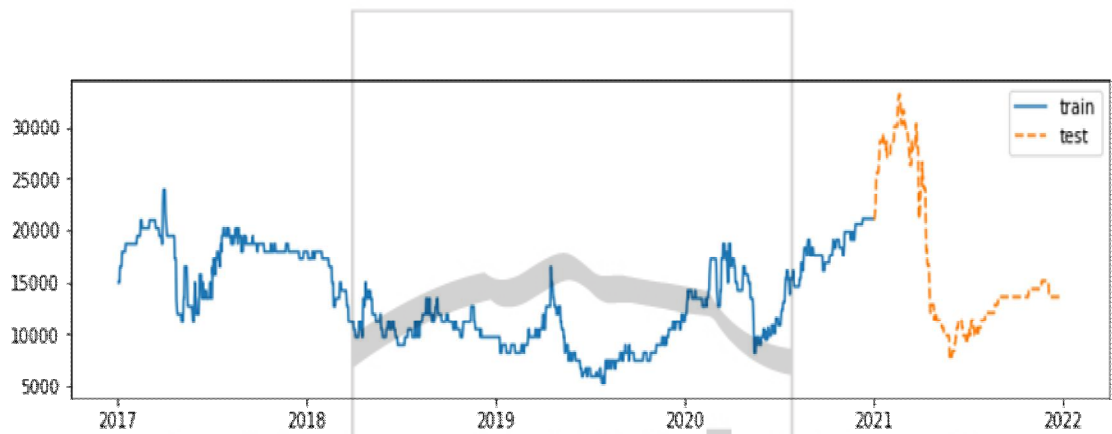
[그림 3-14] 평균 풍속(m/s) train, test 데이터



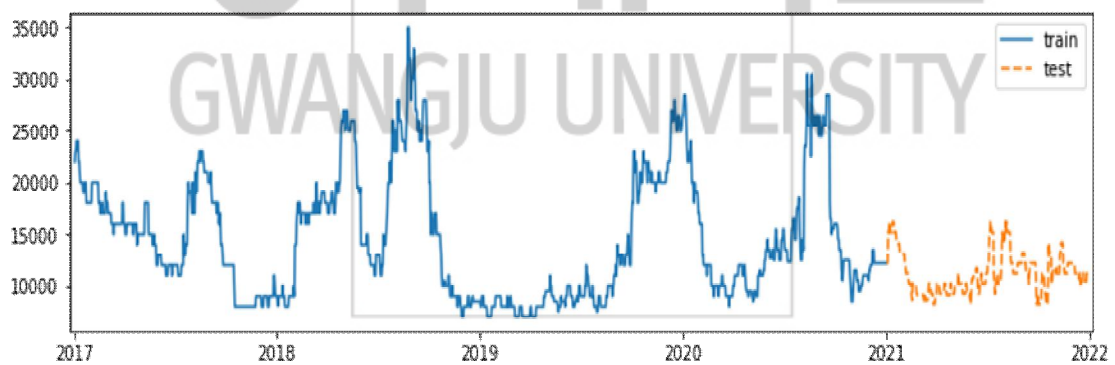
[그림 3-15] 평균 상대습도(%) train, test 데이터



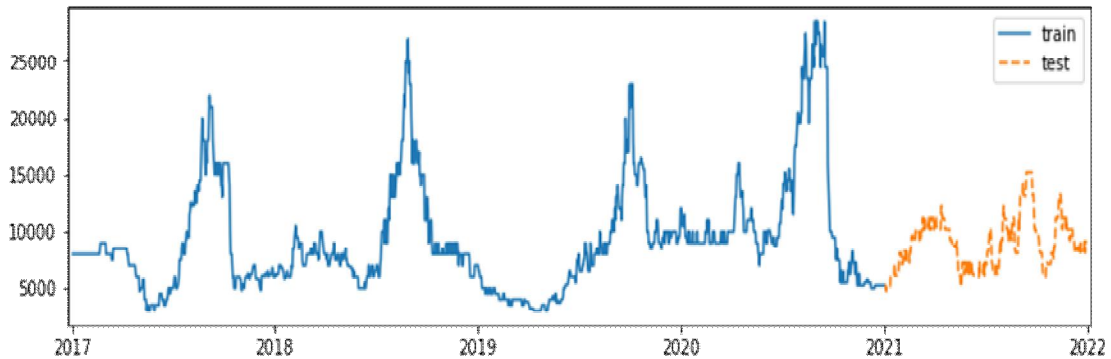
[그림 3-16] 소비자 물가 train, test 데이터



[그림 3-17] 농산물 양과 가격 train, test 데이터



[그림 3-18] 농산물 무 가격 train, test 데이터



[그림 3-19] 농산물 배추 가격 train, test 데이터

이렇게 선택된 데이터를 활용하여 딥러닝 모델 학습을 진행하면, 데이터의 특성을 잘 반영하지 못해 성능이 떨어진다. 농산물 가격예측 성능을 높이기 위해 두 가지 전처리 방법을 진행한다. 첫 번째 방법으로는 데이터 스케일링(Data Scaling)을 사용했다. 각 데이터 속성별 범위가 다를 경우 딥러닝 모델이 잘 동작하지 않아 데이터 스케일링 작업이 필요하다. 여기서는 데이터 속성별로 최솟값과 최댓값을 0부터 1사이로 스케일링하는 Min-Max Scaling 방식을 사용하여 스케일링했다.

	양파	평균기온(°C)	평균 풍속(m/s)	평균 상대습도(%)	물가
2017-01-02	0.520	0.3800	0.112903	0.768309	0.008230
2017-01-03	0.520	0.3425	0.112903	0.591212	0.008230
2017-01-04	0.520	0.3250	0.080645	0.708389	0.008230
2017-01-05	0.600	0.3700	0.209677	0.631158	0.008230
2017-01-06	0.600	0.4050	0.209677	0.559254	0.008230

[그림 3-20] 양파 데이터 셋 스케일링

두 번째 방법으로는 딥러닝 모델 학습 시 시간 순서를 반영할 수 있도록 데이터 구조를 변경했다. 여기서는 5일 전 과거부터 현재까지의 데이터를 보고 미래인 1일 후의 가격을 예측할 수 있도록 구조를 변경했다. 5일 전 과거부터 미래 1일 후

의 가격을 예측하기 위해 데이터는 $t-5$, $t-4$, $t-3$, $t-2$, $t-1$, t , $t+1$ 로 변경하여 모델을 학습했다. 예측하기 위한 데이터 속성들은 $t-5$ 부터 t 까지 구성되어 있고, 예측하고자 하는 1일 후 농산물(양파, 무, 배추) 가격은 $t+1$ 로 구성했다. 여기서 말하는 t 는 현재 시점을 의미하며, $t-1$ 은 현재를 기준으로 1일 전, $t-2$ 는 현재를 기준으로 2일 전으로 $t-5$ 는 5일 전이 된다. 예를 들어 2017년 1월 7일을 t 라고 가정하면 2017년 1월 6일은 $t-1$, 2017년 1월 2일은 $t-5$ 가 된다. 변경된 데이터를 보면 과거 5일이 없는 2017년 1월 2일부터 2017년 1월 6일까지는 삭제 처리했다. $t+1$ 은 1일 후를 의미하며 목표값(label)인 농산물 가격이다. 이와 같이 두 가지 전처리 방법을 활용하여 무와 배추도 똑같이 진행했다. 완료된 데이터는 데이터 크기(data), 시간 스텝(time step), 속성(feature) 이렇게 3차원으로 구성된다.

tensor	target	X														
feature	y	양파							평균기온(°C)			...	평균 상대습도(%)			
time step	t+1	t-5	t-4	t-3	t-2	t-1	t	t-5	t-4	t-3	...	t-3	t-2	t-1	t	
2017-01-07	0.600	0.520	0.520	0.520	0.600	0.600	0.600	0.3800	0.3425	0.3250	...	0.708389	0.631158	0.559254	0.632490	
2017-01-08	0.680	0.520	0.520	0.600	0.600	0.600	0.600	0.3425	0.3250	0.3700	...	0.631158	0.559254	0.632490	0.664447	
2017-01-09	0.680	0.520	0.600	0.600	0.600	0.600	0.680	0.3250	0.3700	0.4050	...	0.559254	0.632490	0.664447	0.555260	
2017-01-10	0.680	0.600	0.600	0.600	0.600	0.680	0.680	0.3700	0.4050	0.3650	...	0.632490	0.664447	0.555260	0.446072	
2017-01-11	0.680	0.600	0.600	0.600	0.680	0.680	0.680	0.4050	0.3650	0.4175	...	0.664447	0.555260	0.446072	0.392810	

[그림 3-21] 양파 데이터 셋 전처리 후

제2절 딥러닝 모델

1. RNN 예측모델

딥러닝 RNN 모델을 사용하여 농산물 가격예측을 수행하였다. 훈련데이터 1,454개, 테스트데이터 359개로 학습하였다. 모델은 농산물 가격예측을 목적으로 회귀분석 방법을 사용한다. RNN은 데이터 개수에 비해 유닛 수가 많으면 파라미터 수가 증가하게 되고, 과대 적합(over fitting)이 발생하게 되어 훈련데이터는 성능이 높지만 예측하고자 하는 테스트데이터는 성능이 떨어진다.[10] [표 3-4]은 RNN 모델 구성이며, [표 3-5]는 사용한 하이퍼 파라미터이다.

[표 3-4] RNN 모델 구성도

Layer	Output Shape	Param
simple_rnn	(None, 5)	55
repeat_vector	(None, 1, 5)	0
simple_rnn	(None, 1, 5)	55
time_distributed	(None, 1, 1)	6
flatten	(None, 1)	0
dense	(None, 1)	2
total params	-	118

[표 3-5] RNN 모델 하이퍼 파라미터

Parameter	Description
input shape	time step : 6, feature : 5
deep learning model	RNN
optimizer	RMSProp
learning rate	0.001
batch size	32
epochs	300
loss function	RMSE
earlystopping_patience	15

1,454개의 훈련데이터를 학습시켜 만들어진 RNN 모델은 359개의 테스트데이터를 사용하여 성능 검증하였다. 학습은 케라스(kearas)를 활용하여 RNN 모델을 사용하였고, 학습 도중 Optimizer로 인해 성능이 미세하게 달라지는 것을 확인하였다. 이에 따라 5번을 반복하여 학습하고, 학습 시 ModelCheckpoint를 사용하여 RMSE가 가장 적을 때 모델을 저장하였다. 학습은 양파, 무, 배추별로 각각 진행하였다. 각각 저장된 모델을 불러와 테스트데이터로 검증한 결과 RMSE 값은 양파 2,059.8076, 무 973.2566, 배추 942.7267으로 나타났다.

[표 3-6] 양파, 무, 배추 RNN 모델 오차

Agricultural products	RMSE
양파	2,059.8076
무	973.2566
배추	942.7267

학습된 모델의 성능을 검증하기 위해 2021년 1월 6일부터 2021년 12월 30일까지 359일간의 농산물 가격을 예측하고 실제 가격과 비교하였다. [표 3-7], [표 3-8], [표 3-9]는 RNN 모델로 양파, 무, 배추 가격을 예측한 결과이며, 평균 오차율은 9.496%, 7.112%, 8.454%로 나타났다.

[표 3-7] RNN 모델 양파 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	25,650	23,497.516081	8.391750
2021.01.07	25,650	24,058.975071	6.202826
2021.01.08	25,650	24,129.320472	5.928575
2021.01.09	25,650	24,098.962337	6.046930
2021.01.10	26,400	23,885.100871	9.526133
...			
2021.12.26	13,650	15,195.956245	11.325687
2021.12.27	13,650	15,486.854106	13.456807
2021.12.28	13,650	15,513.040289	13.648647
2021.12.29	13,650	15,475.249082	13.371788
2021.12.30	13,650	15,428.809986	13.031575

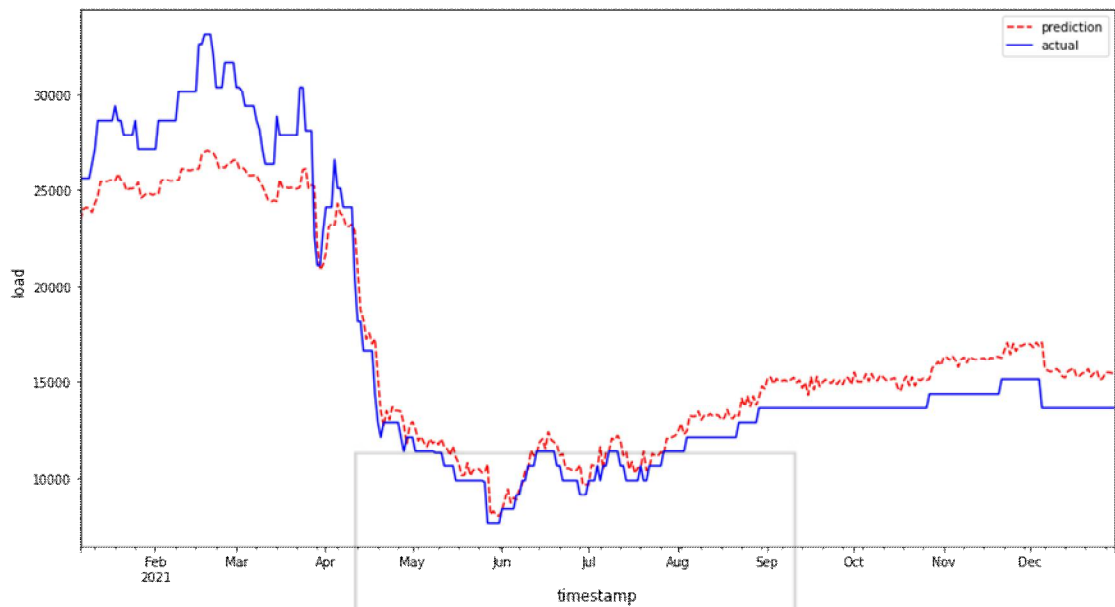
[표 3-8] RNN 모델 무 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	16,200	16,577.274442	2.328855
2021.01.07	15,200	16,390.012145	7.829027
2021.01.08	15,200	15,398.236275	1.304186
2021.01.09	15,200	15,300.912619	0.663899
2021.01.10	15,200	15,220.450163	0.134541
...			
2021.12.26	11,200	8,778.530568	21.620263
2021.12.27	11,200	10,119.446397	9.647800
2021.12.28	11,200	9,787.882984	12.608188
2021.12.29	11,200	10,217.652500	8.770960
2021.12.30	11,200	9,900.282502	11.604621

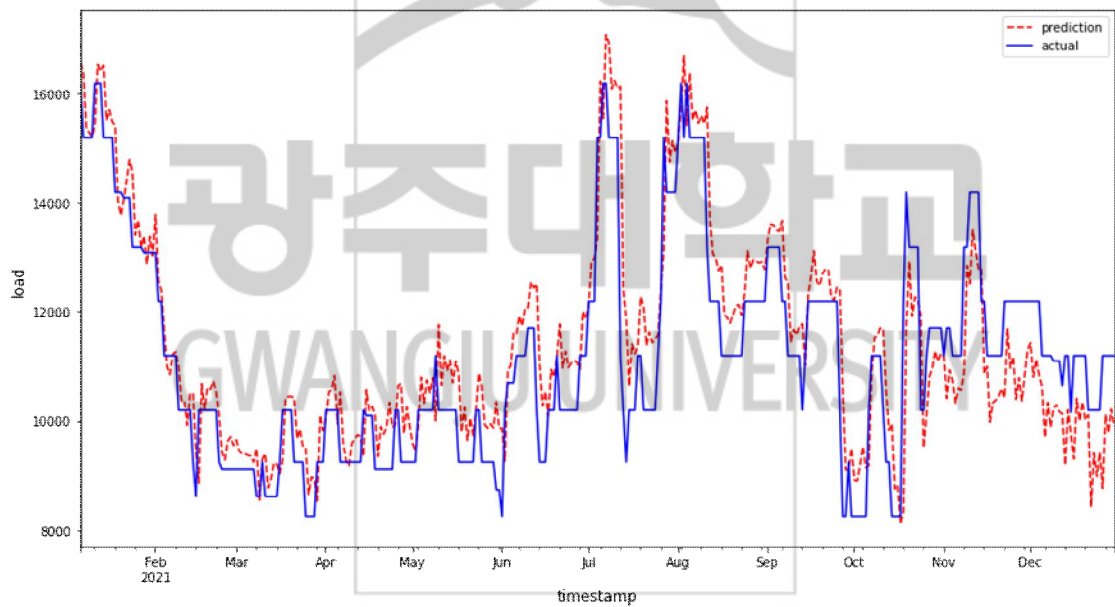
[표 3-9] RNN 모델 배추 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	5,120	4,685.374483	8.488780
2021.01.07	5,120	4,638.828691	9.397877
2021.01.08	5,120	4,780.317329	6.634427
2021.01.09	5,120	4,700.561129	8.192165
2021.01.10	5,120	4,611.821830	9.925355
...			
2021.12.26	9,120	7,062.381774	22.561603
2021.12.27	8,120	7,830.745004	3.562254
2021.12.28	9,120	7,224.983409	20.778691
2021.12.29	9,120	8,401.592933	7.877270
2021.12.30	9,120	8,271.205209	9.306960

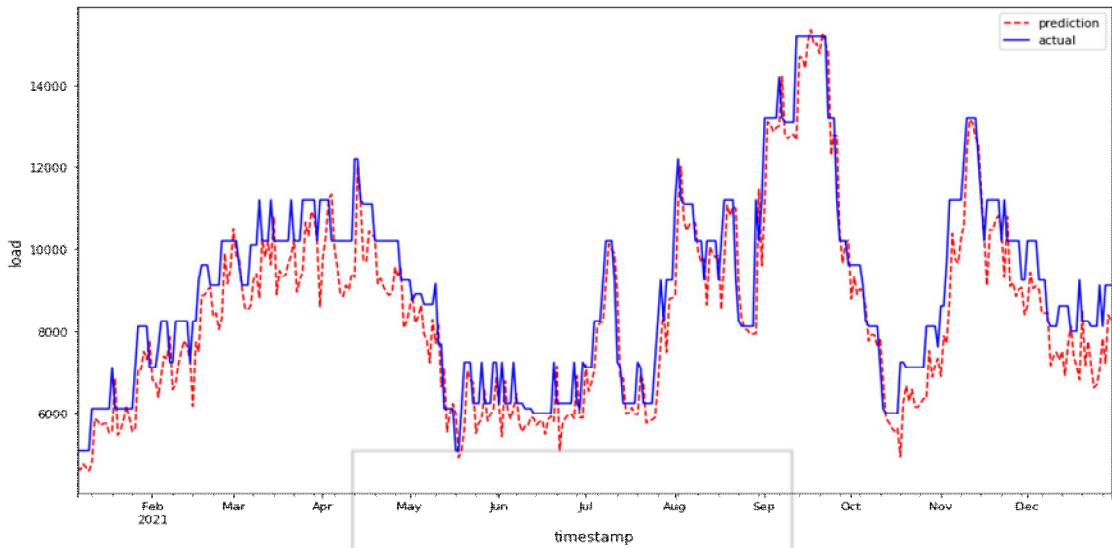
[그림 3-22], [그림 3-23], [그림 3-24]는 RNN 모델을 사용한 양파, 무, 배추 가격예측 결과를 시각화한 것이다. 파란색 선은 실제 가격이며, 빨간색 선은 예측값이다. 양파 예측 결과는 1월, 2월 3월까지의 실제 가격보다 낮은 가격으로 예측하고, 8월 이후로는 실제 가격보다 높은 가격으로 예측하여 오차가 심한 모습을 보인다. 무의 예측 결과는 전반적으로 높은 성능을 가지지만, 1월부터 10월까지의 실제 가격보다 높게 예측되고 11월, 12월은 실제 가격보다 낮게 예측한 모습을 보인다. 배추의 예측 결과는 전반적으로 높은 성능을 가지지만, 실제 가격보다 낮은 가격으로 예측한 모습을 보인다.



[그림 3-22] RNN 모델 양과 가격예측 결과



[그림 3-23] RNN 모델 무 가격예측 결과



[그림 3-24] RNN 모델 배추 가격예측 결과

2. LSTM 예측모델

딥러닝 LSTM 모델을 사용하여 농산물 가격예측을 수행하였다. 훈련데이터 1,454개, 테스트데이터 359개로 학습하였다. 모델은 농산물 가격예측을 목적으로 회귀분석 방법을 사용한다. LSTM은 RNN의 단점을 개선하여 예측 성능이 좋다고 알려져 있다. [표 3-10]은 LSTM 모델 구성이며, [표 3-11]은 사용한 하이퍼 파라미터이다.

[표 3-10] LSTM 모델 구성도

Layer	Output Shape	Param
lstm	(None, 5)	220
repeat_vector	(None, 1, 5)	0
lstm	(None, 1, 5)	220
time_distributed	(None, 1, 1)	6
flatten	(None, 1)	0
dense	(None, 1)	2
total params	-	448

[표 3-11] LSTM 모델 하이퍼 파라미터

Parameter	Description
input shape	time step : 6, feature : 5
deep learning model	LSTM
optimizer	RMSProp
learning rate	0.001
batch size	32
epochs	300
loss function	RMSE
earlystopping_patience	15

1,454개의 훈련데이터를 학습시켜 만들어진 LSTM 모델은 359개의 테스트데이터를 사용하여 성능 검증하였다. 학습은 케라스(kearas)를 활용하여 LSTM 모델을 사용하였고, 학습 도중 Optimizer로 인해 성능이 미세하게 달라지는 것을 확인하였다. 이에 따라 5번을 반복하여 학습하고, 학습 시 ModelCheckpoint를 사용하여 RMSE가 가장 적을 때 모델을 저장하였다. 학습은 양파, 무, 배추별로 각각 진행하였다. 각각 저장된 모델을 불러와 테스트데이터로 검증한 결과 RMSE 값은 양파 1,561.6134, 무 717.8054, 배추 750.0324으로 나타났다.

[표 3-12] 양파, 무, 배추 LSTM 모델 오차

Agricultural products	RMSE
양파	1,561.6134
무	717.8054
배추	750.0324

학습된 모델의 성능을 검증하기 위해 2021년 1월 6일부터 2021년 12월 30일까지 359일간의 농산물 가격을 예측하고 실제 가격과 비교하였다. [표 3-13], [표 3-14], [표 3-15]는 LSTM 모델로 양파, 무, 배추 가격을 예측한 결과이며, 평균 오차율은 5.153%, 4.51%, 6.735%로 나타났다.

[표 3-13] LSTM 모델 양파 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	25,650	23,070.366472	10.057051
2021.01.07	25,650	23,596.220255	8.006939
2021.01.08	25,650	23,745.051563	7.426700
2021.01.09	25,650	23,844.324589	7.039670
2021.01.10	26,400	23,714.452028	10.172530
...			
2021.12.26	13,650	14,568.174608	6.726554
2021.12.27	13,650	14,636.026114	7.223635
2021.12.28	13,650	14,695.542097	7.659649
2021.12.29	13,650	14,644.857287	7.288332
2021.12.30	13,650	14,587.458014	6.867824

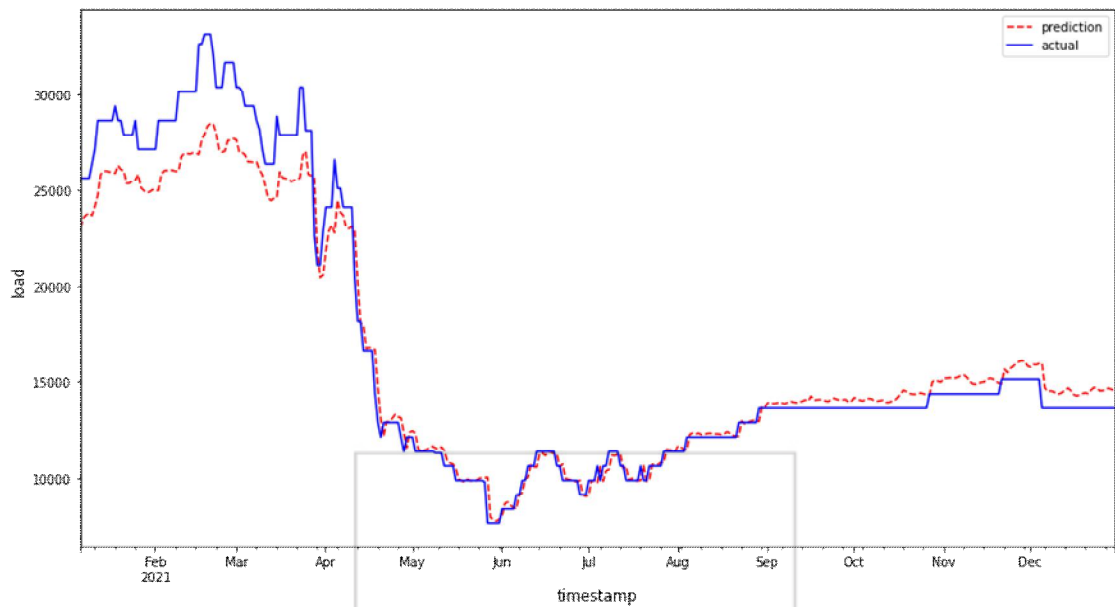
[표 3-14] LSTM 모델 무 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	16,200	15,707.250357	3.041664
2021.01.07	15,200	15,892.627597	4.556761
2021.01.08	15,200	14,905.307770	1.938765
2021.01.09	15,200	14,717.799306	3.172373
2021.01.10	15,200	14,748.636961	2.969494
...			
2021.12.26	11,200	11,036.055744	1.463788
2021.12.27	11,200	11,347.453475	1.316549
2021.12.28	11,200	11,277.884960	0.695401
2021.12.29	11,200	11,364.619672	1.469819
2021.12.30	11,200	11,481.424749	2.512721

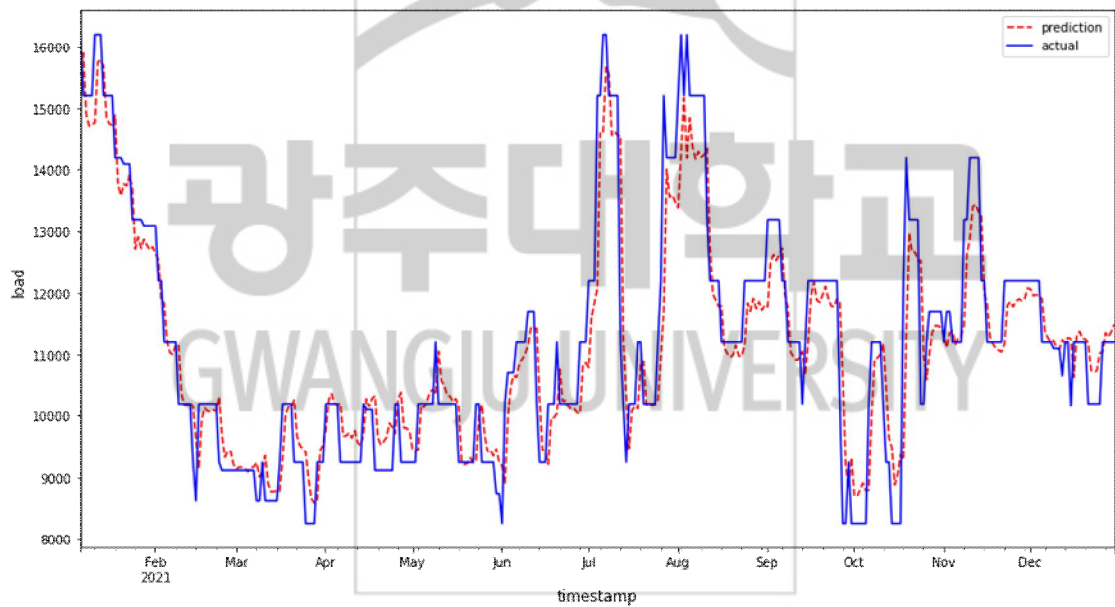
[표 3-15] LSTM 모델 배추 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	5,120	5,281.324625	3.150872
2021.01.07	5,120	5,599.095304	9.357330
2021.01.08	5,120	5,664.314814	10.631149
2021.01.09	5,120	5,493.447863	7.293904
2021.01.10	5,120	5,424.042307	5.938326
...			
2021.12.26	9,120	8,756.772034	3.982763
2021.12.27	8,120	9,449.179620	16.369207
2021.12.28	9,120	8,467.528895	7.154288
2021.12.29	9,120	9,025.951870	1.031229
2021.12.30	9,120	9,253.186516	1.460378

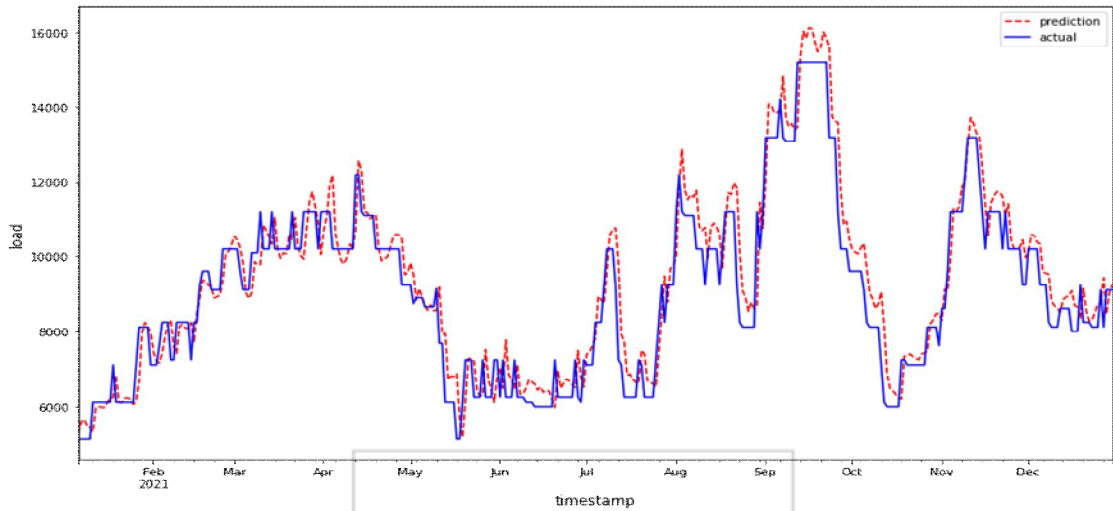
[그림 3-25], [그림 3-26], [그림 3-27]은 LSTM 모델을 사용한 양파, 무, 배추 가격예측 결과를 시각화한 것이다. 파란색 선은 실제 가격이며, 빨간색 선은 예측 값이다. 양파 예측 결과는 1월, 2월, 3월까지의 실제 가격보다 낮은 가격으로 예측 되는 모습이나 4월 이후에는 오차 없이 좋은 성능을 보인다. 무의 예측 결과는 8월에 가격이 급격한 변화를 보일 때 오차가 가장 큰 것으로 나타나며 전반적으로 좋은 성능을 나타낸다. 배추의 예측 결과는 무와 마찬가지로 8월에 가격이 급격한 변화를 보일 때 오차가 있으나 전반적으로 좋은 성능을 나타낸다.



[그림 3-25] LSTM 모델 양과 가격예측 결과



[그림 3-26] LSTM 모델 무 가격예측 결과



[그림 3-27] LSTM 모델 배추 가격예측 결과

3. GRU 예측모델

딥러닝 GRU 모델을 사용하여 농산물 가격예측을 수행하였다. 훈련데이터 1,454개, 테스트데이터 359개로 학습하였다. 모델은 농산물 가격예측을 목적으로 회귀분석 방법을 사용한다. GRU는 LSTM보다 학습속도 및 예측 속도가 빠르다는 장점이 있으나 LSTM보다 성능이 비슷하거나 떨어진다고 알려져 있다. [표 3-16]은 GRU 모델 구성이며, [표 3-17]은 사용한 하이퍼 파라미터이다.

[표 3-16] GRU 모델 구성도

Layer	Output Shape	Param
gru	(None, 5)	180
repeat_vector	(None, 1, 5)	0
gru	(None, 1, 5)	180
time_distributed	(None, 1, 1)	6
flatten	(None, 1)	0
dense	(None, 1)	2
total params	-	368

[표 3-17] GRU 모델 하이퍼 파라미터

Parameter	Description
input shape	time step : 6, feature : 5
deep learning model	GRU
optimizer	RMSProp
learning rate	0.001
batch size	32
epochs	300
loss function	RMSE
earlystopping_patience	15

1,454개의 훈련데이터를 학습시켜 만들어진 GRU 모델은 359개의 테스트데이터를 사용하여 성능 검증하였다. 학습은 케라스(kearas)를 활용하여 GRU 모델을 사용하였고, 학습 도중 Optimizer로 인해 성능이 미세하게 달라지는 것을 확인하였다. 이에 따라 5번을 반복하여 학습하고, 학습 시 ModelCheckpoint를 사용하여 RMSE가 가장 적을 때 모델을 저장하였다. 학습은 양파, 무, 배추별로 각각 진행하였다. 각각 저장된 모델을 불러와 테스트데이터로 검증한 결과 RMSE 값은 양파 1,711.3141, 무 717.8054, 배추 761.3761로 나타났다.

[표 3-18] 양파, 무, 배추 GRU 모델 오차

Agricultural products	RMSE
양파	1,711.3141
무	717.8054
배추	761.3761

학습된 모델의 성능을 검증하기 위해 2021년 1월 6일부터 2021년 12월 30일까지 359일간의 농산물 가격을 예측하고 실제 가격과 비교하였다. [표 3-19], [표 3-20], [표 3-21]은 GRU 모델로 양파, 무, 배추 가격을 예측한 결과이며, 평균 오차율은 6.863%, 4.51%, 7.416%로 나타났다.

[표 3-19] GRU 모델 양파 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	25,650	23,370.399490	8.887331
2021.01.07	25,650	24,098.394603	6.049144
2021.01.08	25,650	24,147.338212	5.858331
2021.01.09	25,650	24,091.952831	6.074258
2021.01.10	26,400	24,099.051744	8.715713
...			
2021.12.26	13,650	13,516.108856	0.980888
2021.12.27	13,650	13,392.004721	1.890075
2021.12.28	13,650	12,998.126239	4.775632
2021.12.29	13,650	13,030.093700	4.541438
2021.12.30	13,650	13,191.353694	3.360046

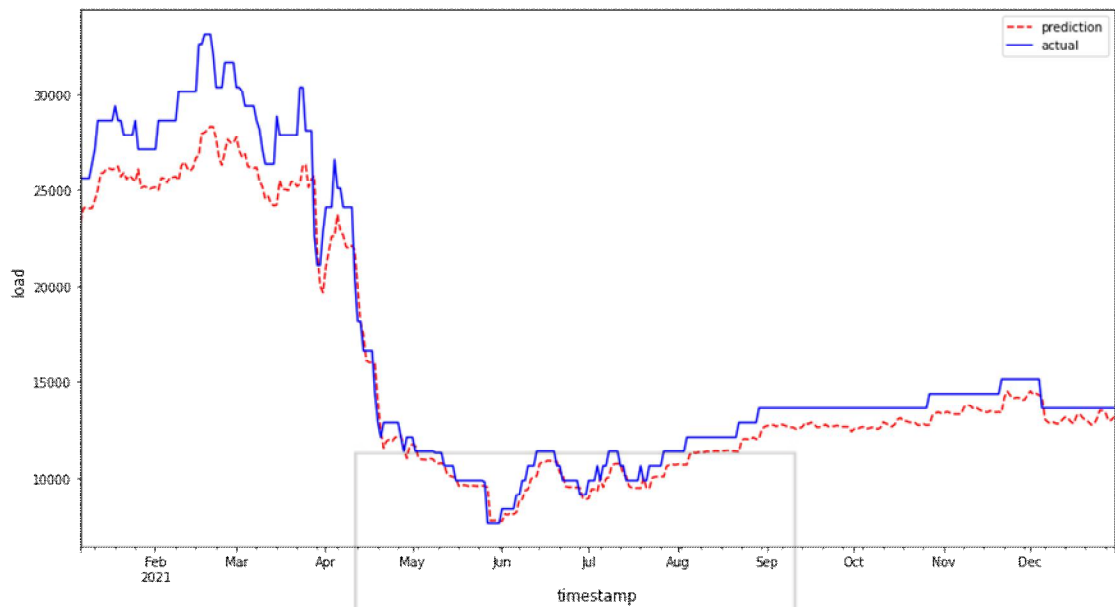
[표 3-20] GRU 모델 무 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	16,200	15,707.250357	3.041664
2021.01.07	15,200	15,892.627597	4.556761
2021.01.08	15,200	14,905.307770	1.938765
2021.01.09	15,200	14,717.799306	3.172373
2021.01.10	15,200	14,748.636961	2.969494
...			
2021.12.26	11,200	11,036.055744	1.463788
2021.12.27	11,200	11,347.453475	1.316549
2021.12.28	11,200	11,277.884960	0.695401
2021.12.29	11,200	11,364.619672	1.469819
2021.12.30	11,200	11,481.424749	2.512721

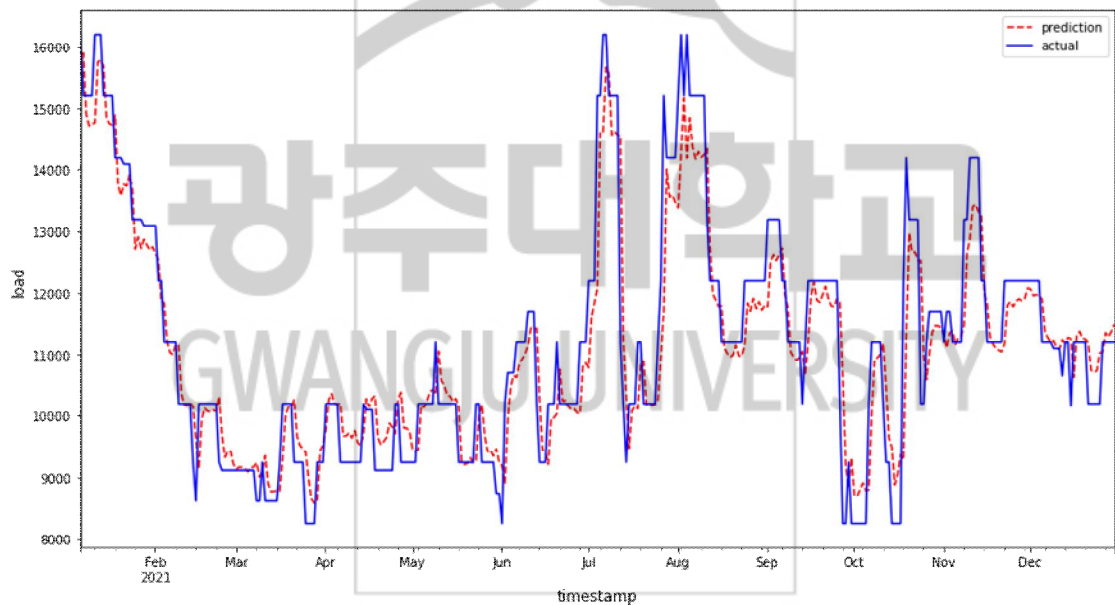
[표 3-21] GRU 모델 배추 가격예측 결과

Date	Label	Predict	Error Rate(%)
2021.01.06	5,120	5,842.198659	14.105443
2021.01.07	5,120	6,052.757710	18.217924
2021.01.08	5,120	6,099.952508	19.139697
2021.01.09	5,120	5,922.378536	15.671456
2021.01.10	5,120	5,869.085826	14.630583
...			
2021.12.26	9,120	9,220.151849	1.098156
2021.12.27	8,120	9,586.946547	18.065844
2021.12.28	9,120	9,001.959577	1.294303
2021.12.29	9,120	9,778.113529	7.216157
2021.12.30	9,120	9,953.135937	9.135262

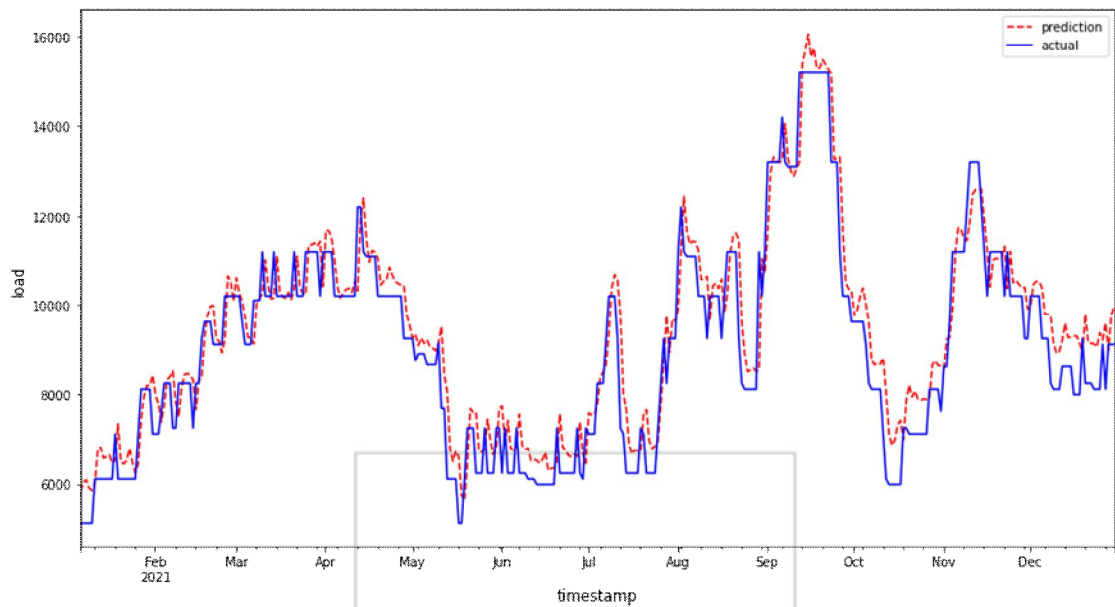
[그림 3-28], [그림 3-29], [그림 3-30]은 GRU 모델을 사용한 양파, 무, 배추 가격예측 결과를 시각화한 것이다. 파란색 선은 실제 가격이며, 빨간색 선은 예측값이다. 양파 예측 결과는 5월, 6월, 7월은 좋은 성능을 보이는 모습이나 그 외에는 실제 가격보다 낮은 가격으로 예측되는 모습이다. 무의 예측 결과는 8월에 가격이 급격한 변화를 보일 때 오차가 가장 큰 것으로 나타나며 전반적으로 좋은 성능을 나타낸다. 배추의 예측 결과는 실제 가격보다 높게 예측되지만, 전반적으로 좋은 성능을 보인다.



[그림 3-28] GRU 모델 양파 가격예측 결과



[그림 3-29] GRU 모델 무 가격예측 결과



[그림 3-30] GRU 모델 배추 가격예측 결과

광주대학교
GWANGJU UNIVERSITY

제4장 실험 평가 방법 및 결과

제1절 실험 평가 방법

농산물 가격예측은 회귀분석으로 회귀분석은 평균 제곱 오차(MSE, Mean Squared Error), 평균 제곱근 오차(RMSE, Root Mean Square Error), 평균 절대 오차(MAE, Mean Absolute Error) 등의 손실함수를 활용하여 성능을 평가할 수 있다. 본 논문에서는 손실함수 중 가장 많이 사용하는 RMSE로 오차를 계산하고, 추가로 백분율 오차율, 정확도를 통해서 성능을 비교하였다. RMSE의 수식은 식 (3)과 같으며, 데이터별로 예측값에서 실제값을 뺀 결과를 전부 더하여 제곱근을 구한 후 평균을 구한 오차를 나타낸다. RMSE는 예측값과 실제값의 차이를 나타내어 0에 가까울수록 높은 성능을 보인다고 할 수 있다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2} \quad (3)$$

MSE, RMSE, MAE 등의 결과는 실제 결과와의 차이를 알 수 있지만, 분류분석 평가지표인 정확도(Accuracy), 재현율(Recall), 정밀도(Precision)처럼 정확한 성능을 알기에는 어려움이 있다. 이에 식 (4)를 통해 백분율 오차율을 구하고 식 (5)를 통해 정확도를 계산하여 성능을 확인했다.

$$\text{백분율오차율} = \frac{|\text{예측값} - \text{실제값}|}{\text{실제값}} \times 100 \quad (4)$$

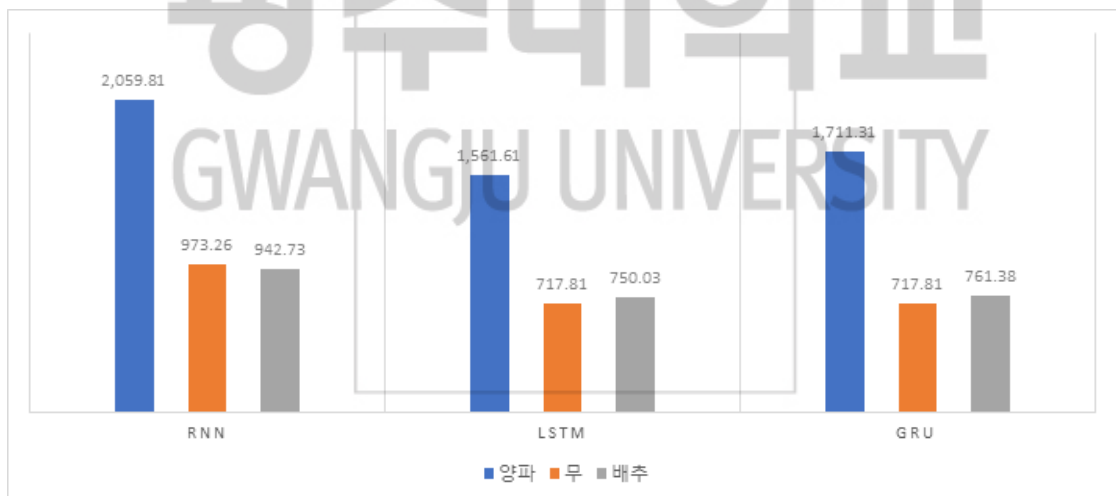
$$\text{정확도}(\%) = 100 - \text{백분율오차율} \quad (5)$$

제2절 실험 결과

최종 실험 결과 양파, 무, 배추 가격예측 모델별 RMSE는 양파 RNN 모델 2,059.8076, 무 RNN 모델 973.2566, 배추 RNN 모델 942.7267, 양파 LSTM 모델 1,561.6134, 무 LSTM 모델 717.8054, 배추 LSTM 모델 750.0324, 양파 GRU 모델 1,711.3141, 무 GRU 모델 717.8054, 배추 GRU 모델 761.3761으로 나타났다. 이 결과 LSTM 모델이 RNN과 GRU보다 우수한 성능을 갖는 것을 알 수 있다. [표 4-1]과 [그림 4-1]은 모델별 RMSE와 그래프를 나타낸다.

[표 4-1] 농산물 가격예측 모델별 RMSE

농산물		RNN	LSTM	GRU
양파	RMSE	2,059.81	1,561.61	1,711.31
무	RMSE	973.26	717.81	717.81
배추	RMSE	942.73	750.03	761.38

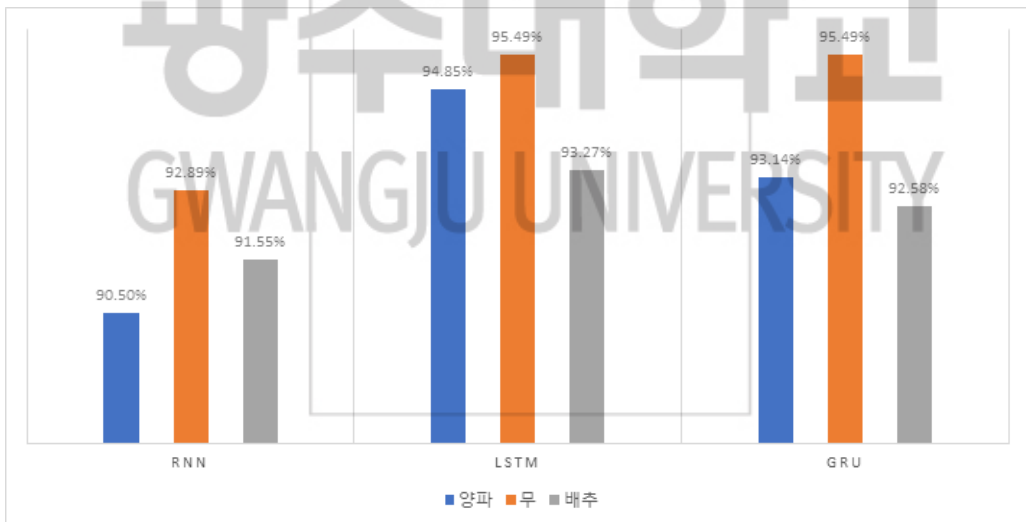


[그림 4-1] 농산물 가격예측 모델별 RMSE 그래프

양파, 무, 배추 가격예측 모델별 정확도(%)는 양파 RNN 모델 90.504%, 무 RNN 모델 92.888%, 배추 RNN 모델 91.546%, 양파 LSTM 모델 94.847%, 무 LSTM 모델 95.49%, 배추 LSTM 모델 93.265%, 양파 GRU 모델 93.137%, 무 GRU 모델 95.49%, 배추 GRU 모델 92.584%으로 나타났다. 이 결과는 이전 시계열 데이터 예측모델과 회귀분석 모델이 보여주는 90% 수준의 정확도보다 우수한 성능을 보인다. [표 4-2]와 [그림 4-2]는 모델별 정확도와 그래프를 나타낸다.

[표 4-2] 농산물 가격예측 모델별 정확도

농산물		RNN	LSTM	GRU
양파	오차율(%)	9.496	5.153	6.863
	정확도(%)	90.504	94.847	93.137
무	오차율(%)	7.112	4.51	4.51
	정확도(%)	92.888	95.49	95.49
배추	오차율(%)	8.454	6.735	7.416
	정확도(%)	91.546	93.265	92.584



[그림 4-2] 농산물 가격예측 모델별 정확도 그래프

제5장 결론 및 향후 연구

본 논문은 농산물 가격예측에 가장 적합한 딥러닝 모델을 찾기 위해 RNN, LSTM, GRU 모델의 성능을 비교하고 평가하였다. 데이터는 기상청 날씨 데이터, 물가 데이터, 농산물 가격 데이터를 사용하고 1,825건의 데이터를 딥러닝 모델로 학습시켜 성능을 평가하고 검증하였다.

실험에 사용한 입력데이터는 2017년부터 2021년까지 총 5년간의 광주 기상정보데이터와 국가통계포털(KOSIS)의 소비자 물가지수 데이터를 사용하였다. 성능향상을 위해 전처리 작업 후 2017년부터 2020년까지 4년간의 훈련데이터, 2021년 1년간의 테스트 데이터로 구성하여 학습하였다. 그 결과 양파, 무, 배추의 예측 정확도는 평균 RNN 91.7%, LSTM 94.6%, GRU 93.7%로 LSTM 모델이 가장 우수한 성능을 보이며 이전 시계열 데이터 예측모델과 회귀분석 모델이 보여주는 90% 수준의 정확도보다 우수한 성능을 보여주었다.

국내 농산물 생산액은 2020년 기준 50조원에 달하며 도·소매시장에서 거래되는 금액은 100조원을 훌쩍 넘는다. 이러한 큰 시장에 아직도 표준화된 데이터와 가격 기준, 시스템이 부족한 상황이며 표준화된 가격 없이 매해 폭락하거나 급등하는 농산물 가격으로 농민들과 중간 구매자들이 고스란히 피해를 보고 있다. 본 연구를 통해 광주 지역의 농산물 가격을 예측하였으며 이를 통하여 농산물의 건전한 유통질서 확립, 수급조절을 유도하여 실효성 있는 가격 안정 대책추진, 소비자와 생산자 입장에서 전반적인 후생 증가가 가능할 것으로 생각된다. 향후 농산물 가격 안정을 위해 양파, 무, 배추뿐만 아니라 다른 농산물도 가격을 예측하고, 오차를 더 줄일 수 있는 다양한 연구가 필요할 것으로 보인다. 또한, 최근의 우크라이나 전쟁으로 인한 곡물가 급등에서 볼 수 있듯이 다양한 외생변수에 대비할 수 있도록 데이터를 수집, 보완하며 수입 의존도가 높은 곡물에 대한 가격예측에도 활용 가능하도록 보완하고자 한다.

참고 문헌

- [1] 남국현, 최영찬, “양파 출하시기 도매가격예측모형 연구”, 농촌지도와 개발, Vol. 22, No. 4, pp423-434, 2015
- [2] 이도영, 양예원, 이주형, 박지홍, 강민구, “Lasso 회귀분석을 활용한 농산물 가격예측 모델 변수 선정 연구”, Vol.4, No.2, pp.123-136, 2021
- [3] 신성호, 이미경, 송사광, “LSTM 네트워크를 활용한 농산물 가격예측 모델”, 한국과학기술정보연구원, Vol.18, No.11, pp.416-429, 2018
- [4] 배경태, “인공신경망 기법을 이용한 최적의 농산물 가격예측모델 개발”, 숭실대학교, 석사학위논문, 2017
- [5] Schmidhuber, Jürgen. "Habilitation thesis: System modeling and optimization." Page 150 ff demonstrates credit assignment across the equivalent of 1,200 layers in an unfolded RNN 1993
- [6] 안성만, 정여진, 이재준, 양지현, “한국어 음소 단위 LSTM 언어모델을 이용한 문장 생성”, 한국지능정보시스템학회, Vol.23, No.2, pp.71-88, 2017
- [7] Hochreiter, Sepp. "Untersuchungen zu dynamischen neuronalen Netzen." Diploma, Technische Universität München 91.1 1991
- [8] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 2014
- [9] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 2014
- [10] 이금성, “전력수요예측을 위한 딥러닝 모델의 성능 비교 연구”, 광주대학교, 박사학위논문, 2020
- [11] https://ko.wikipedia.org/wiki/순환_신경망
- [12] <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>