



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](#)

석 사 학 위 논 문

감귤 출하량과 가격에 대한 예측 모형
개발 및 영향요인 분석

김 성 율

제주대학교 대학원

전산통계학과

2023년 08월

감귤 출하량과 가격에 대한 예측 모형 개발 및 영향요인 분석

이 논문을 [전산통계학] 석사 학위논문으로 제출함

김 성 율

제주대학교 대학원

전산통계학과

지도 교수 서 윤 암

김성율의 전산통계학 석사 학위논문을 인준함

2023년 06월

심사위원장	오 승 희	인
위 원	서 윤 암	인
위 원	윤 여 찬	인

목 차

1. 서론	5
1.1. 연구 배경	5
1.2. 연구 목적	10
2. 연구 방법	12
2.1. 감귤 출하량 및 가격 예측 모형	12
2.2. 심층학습 모형	15
3. 연구 자료	21
3.1. 활용 데이터	21
3.2. 데이터 가공 방법	26
4. 분석 결과	31
4.1. 탐색적 데이터 분석	31
4.2. 실험 모형 및 예측 결과	50
5. 결론	76
5.1. 요약	76
5.2. 연구의 한계 및 향후 과제	77
5.3. 기대 효과	78
6. 참고문헌	79

감귤 출하량과 가격에 대한 예측 모형 개발 및 영향요인 분석

김 성 울

제주대학교 대학원 전산통계학과

요 약

제주도의 1차 산업 규모는 4조2,249억이며, 그중 농산물의 수입은 40.3%인 1조 7,005억이다. 사면이 바다인 지역 특색에 비해 농업 비중이 높으며 농업 중에도 감귤로 인한 수익은 9,508억원(55.92%)으로 농산물 수입의 절반 이상을 차지한다. 그러므로 제주에서 감귤 생산, 공급의 안정화는 생산자 이외에도 유통업자, 소비자, 정부, 학계 등 다양한 이해 관계자에게 중요한 주제이다. 기존 농산물에 관한 연구 현황을 살펴보면 전반적인 농작물의 생산이나 다른 과육의 생산성 관련 연구논문은 다수 있으나 제주 밀감의 대부분을 차지하는 노지온주와 하우스온주에 대해 인공지능을 활용하여 생산량이나 가격을 예측하는 분석 연구는 부재하다. 이에 본 연구에서는 작물 생산과 가격 예측 논문의 기법을 두루 참고하여 제주 감귤 맞춤형으로 개선하고, 노지온주와 하우스온주의 출하량 및 가격 예측 모형을 개발하려고 한다. 감귤 출하량 및 가격 예측에는 다양한 요인 변수가 활용되며 시장의 변동성, 재배 환경(기온, 습도, 일사량, 강우량, 적설량 등), 경쟁 과일의 수요 및 공급 등 다양한 요인 등이 해당한다. 본 연구에서는 과일 출하량

및 가격 변화와 상관관계가 있는 다양한 국·내외 영향요인을 분석하고, 시간에 따른 각 요인을 독립변수로 하여 출하량과 가격을 예측하는 모형으로 설계했다. 이때 활용되는 국내 변동 요인으로는 감귤 출하량, 가격 현황, 기상 정보 및 공휴일 등이 있으며, 국외 변동 요인으로는 경쟁 과일의 수입 과일 데이터를 사용한다. 최적의 예측 모델을 선정하기 위해 다양한 시계열 모델들을 비교 분석하였다. 본 연구에서는 예측 정확도 향상을 위해 심층학습 기반의 LSTM, Seq2Seq, Attention 모델들을 활용하여 감귤 출하량 및 가격 예측을 수행하고, 예측 정확도와 예측 안정성을 평가하였다. 본 연구의 결과는 감귤 출하량 및 가격 예측에 영향을 미치는 요인 변수의 중요성을 확인하고, 다양한 시계열 모델의 윈도우 사이즈를 조절하여 시나리오별 출하량과 가격을 시뮬레이션해 볼 수 있을 것으로 기대된다. 심층학습 기반으로 출하량 및 가격 예측 모델들의 결과를 시뮬레이션해 비교 제시함으로써, 농산물 시장의 효율적인 운영과 농산물 생산자 및 소비자의 이익 극대화 관련 연구 및 실무 분야에 도움이 될 것으로 기대된다.

1. 서론

1.1. 연구 배경

1.1.1. 농산물 시장 현황 및 지역별 특산물

농산물 시장은 경제적인 중요성을 가지고 있다. 한국 농산물 시장은 1985년 국내 최초 농수산물 공영도매시장인 가락동 농수산물 도매시장을 시작으로 공영도매시장 중심의 농산물 유통체계가 형성되어 도매시장의 거래량을 통해 그 규모를 파악할 수 있다. 특히 청과부류의 경우 국내 청과물 생산량의 50% 이상이 공영도매시장을 통해 유통되고 있으며, [표1]을 보면 가락시장은 대표적 공영도매시장으로 전체 공영도매시장 청과물 유통의 약 35%를 차지할 정도로 농산물 유통에서 큰 비중을 차지하고 있다. (정소영, 2022)

[표 1] 가락시장 청과부류 거래물량 및 거래금액 추이

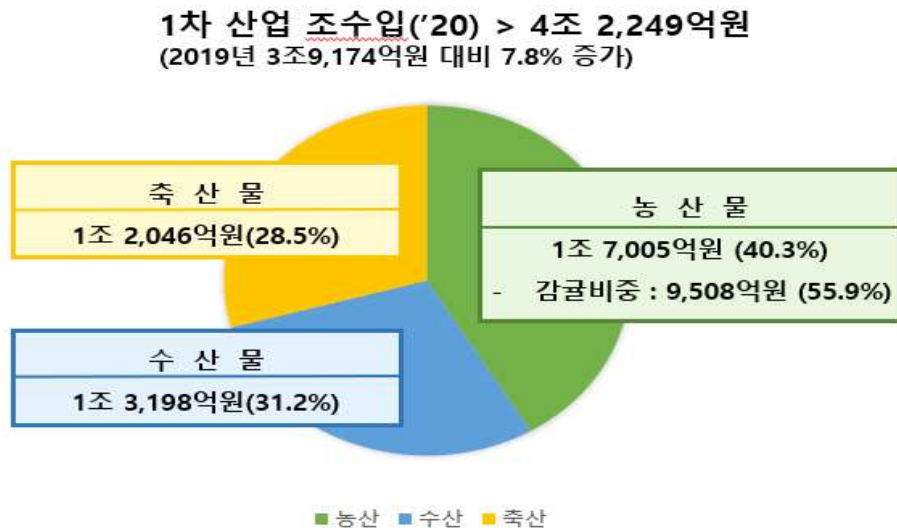
구분	거래물량(천톤)		거래금액(억원)	
	가락시장	전국 공영도매시장	가락시장	전국 공영도매시장
2016년	2,391(34.0%)	7,027	43,230(36.6%)	118,099
2017년	2,410(34.4%)	7,013	41,270(35.8%)	115,155
2018년	2,344(34.3%)	6,843	42,204(36.1%)	117,084
2019년	2,331(33.6%)	6,945	40,292(35.6%)	113,146
2020년	2,281(34.3%)	6,658	45,338(35.9%)	126,318

주: 1) 전국 공영도매시장 33곳 중 청과부류를 거래하는 32곳 대상
주: 2) 괄호 안 수는 전국 공영도매시장 청과부류 유통량 중 가락시장 유통량이 차지하는 비율
출처: 농림축산식품부 도매시장통계연보(2020)

한국은 지역마다 특산물을 보유하고 있으며, 각 지역의 특산물은 지역경제에 큰 영향을 미치고 있다. 예를 들어, 강원도는 고랭지 특성으로 인해 감자와 청정 수산물을 생산하며, 제주도는 기후와 토양 조건이 적합하여 감귤을 생산하고 있다. (김은정, 2007) 또한, 경상지역은 곡류와 과일 등을 생산하고, 전라지역은 쌀과 고추 등을 생산하는 등 각 지역의 특산물은 그 독특한 매력과 경제적 가치를 갖고 있다. 본 연구에서는 제주지역의 특산물인 제주 감귤을 대상으로 연구를 진행하였다.

1.1.2. 제주 감귤 시장

제주특별자치도 친환경농업정책과(2021)에 따르면 제주도의 1차 산업 규모는 4조2,249억이며, 그중 농산물의 수입은 40.3%인 1조 7,005억이다. 사면이 바다인 지역 특색에 비해 농업 비중이 높으며 농업 중에도 감귤로 인한 수익은 9,508억 원(55.92%)으로 농산물 수입의 절반 이상을 차지한다. 그러므로 제주에서 감귤 생산, 공급의 안정화는 생산자 이외에도 유통업자, 소비자, 정부, 학계 등 다양한 이해 관계자에게 중요한 주제이다.



출처: 제주특별자치도 친환경농업정책과, 2021년 제주농축산식품 현황

[그림 1] 제주 농산물 조수입 현황

또한, [표2]를 보면 전국 감귤 재배 면적인 20,860ha(헥타르) 중 제주가 차지하는 면적이 20,748ha(헥타르)로 99.46%에 이른다. 또한, 제주 감귤 생산량은 2022년 기준 634,487톤으로 전국 감귤 공급량(635,835톤)의 99.79%를 담당하고 있을 만큼 감귤 시장은 제주에서 매우 큰 시장이다. (통계청 농어업통계과, 2023)

[표 2] 전국 시도별 감귤 재배 면적 및 생산량

시도별	감귤: 면적(ha)	10ha당 생산량(kg)	생산량(톤)
계	20,860	3,048	635,835
서울특별시	0	0	0
부산광역시	0	996	0
대구광역시	0	0	0
인천광역시	0	0	0
광주광역시	0	0	0

대전광역시	0	0	0
울산광역시	0	996	0
세종특별자치시	0	0	0
경기도	0	0	0
강원도	0	0	0
충청북도	0	1,457	0
충청남도	5	743	35
전라북도	8	520	39
전라남도	91	1,298	1,175
경상북도	4	1,236	51
경상남도	5	996	48
제주도	20,748	3,058	634,487

출처: 통계청, 2022년 농작물생산조사 과일 생산량

하지만 [표3]과 [표4]를 보면 최근 6년간 감귤 생산량과 조수입의 변동 폭이 일정하지 않아 생산량은 약 10만톤 이상, 조수입은 약 2배 이상 차이가 나는 경우도 발생하는 것을 알 수 있다. 이는 농산물 시장의 안정적인 운영과 농산물 생산자의 수익, 소비자의 구매, 유통업체의 재고 관리 전반에 부정적인 영향을 초래한다. (제주특별자치도 친환경농업정책과, 2021)

[표 3] 감귤 품종별 생산량 현황

(단위: ha, 톤)

연도별	구분	합계	온주밀감				만감류		
			소계	노지	하우스	월동	소계	노지	하우스
2020	면적	20,038	16,058	14,752	363	943	3,980	623	3,357
	생산량	654,864	572,486	515,778	25,358	31,350	82,378	4,018	78,360
2019	면적	20,059	16,086	14,815	339	932	3,973	643	3,330
	생산량	631,310	548,868	491,149	27,543	30,176	82,442	3,561	78,881
2018	면적	20,090	16,154	14,898	321	935	3,936	642	3,294
	생산량	607,638	527,475	467,600	22,898	36,977	80,163	3,574	76,589
2017	면적	20,140	16,208	15,006	301	901	3,932	669	3,263
	생산량	576,722	497,903	440,254	22,637	35,012	78,819	3,593	75,226
2016	면적	20,491	18,230	16,610	284	1,336	2,261	101.9	2,159.1
	생산량	599,642	525,727	466,817	21,660	37,250	73,915	2,389	71,526
2015	면적	20,523	18,411	16,775	250	1,386	2,112	99	2,013
	생산량	635,032	567,628	519,243	20,401	27,982	67,406	2,190	65,216

출처: 제주특별자치도 친환경농업정책과, 2021년 제주 농축산식품 현황

[표 4] 감귤 품종별 조수입 현황

(단위: 백만원)

연도별	합계	온주밀감					기타 과수류
		소계	노지	하우스	월동	만감류	
2020	1,008,800	950,828	471,881	100,603	68,819	309,526	58,052
2019	903,904	850,618	403,240	90,703	77,084	279,591	53,286
2018	989,248	940,239	503,034	81,046	89,265	266,894	49,009
2017	998,135	945,792	537,079	80,771	77,277	250,665	52,343
2016	959,005	911,392	491,144	76,087	106,269	237,892	47,613
2015	649,921	602,196	292,363	63,021	47,966	198,846	47,725

출처: 제주특별자치도 친환경농업정책과, 2021년 제주 농축산식품 현황

필자는 본 연구를 통해 이와 같은 문제를 해결하고 감귤 시장의 수요와 공급을 조절할 수 있는 감귤 출하량과 가격 예측 알고리즘을 개발하려 한다. 본 연구에서 개발한 알고리즘은 농산물 시장의 안정적인 운영과 농산물 생산자의 수익 예측, 소비자의 구매 결정, 유통업체의 재고 관리를 위해 중요한 정보로 활용될 수 있을 것이다.

1.1.3. 관련 연구 및 한계점

기존 농산물에 관한 연구 현황을 살펴보면 전반적인 농작물의 생산이나 다른 과육의 생산성 및 가격 관련 연구논문은 다수 있으나 제주 밀감의 대부분을 차지하는 노지온주와 하우스온주에 대해 인공지능을 활용하여 생산량이나 가격을 예측하는 분석 연구는 부재하다.

해외 연구로는 농업대국인 인도의 농산물 출하 및 가격을 연구한 Deshmukh 등(2022), Kuriakose과 Singh(2022), Purohit 등(2021)이 있다.

Deshmukh 등(2022)은 1995~2010년 사이의 작물 수확량 데이터를 사용하여 기계 학습 모델(Random Forest, Decision Tree, KNN, XGBoost)로 수확량을 예측하였다. 모델의 검증은 평균 절대 오차(MAE), 평균 제곱 오차(MSE), R2-Score를 사용하여 63~85%의 정확도를 보였다.

Kuriakose와 Singh (2022)는 기후와 작물의 수확량 간의 높은 상관관계를 가지고 있다는 것을 설명하였다. 또한, 다양한 장소에 대해 딥러닝 계열의 LSTM을 사용하여 미래의 기후 조건과 강우량(단위:mm)을 예측하고, 해당 기후 예측 결과

를 토대로 작물의 수확량을 예측하는 방법을 제시하였다.

Purohit 등(2021)은 농업대국인 인도의 대표 작물인 토마토, 양파, 감자의 월 소매 및 도매가격 데이터를 사용하여 7가지 하이브리드 방법(Additive-ETS-SVM, Additive-ETS-LSTM, Multiplicative-ETS-ANN, Multiplicative-ETS-SVM, Multiplicative-ETS-LSTM, Multiplicative-ARIMA-SVM, Multiplicative-ARIMA-LSTM)으로 작물의 최고 가격을 예측하였다. 모델의 검증은 MAE, SMAE, RMSE를 사용하였다. 시뮬레이션 결과 모든 작물에 대해 최적의 성능을 보이는 모델은 없었으며, 다양한 방법을 실험하고 작물별 서로 다른 모델을 적용해야 최적의 예측 성능을 제공할 수 있었다.

국내 연구로는 농산물의 생산량 및 가격을 예측한 유동완과 박종범(2022), 강민철과 이승호(2019), 김세원과 김영희(2021), 정대호와 조영열(2022)이 있다.

유동완과 박종범(2022)은 농산물 수급 안정을 위해 채소가격 안정제 대상 품목 중 소비자물가지수 가중치가 높은 마늘, 배추, 양파를 대상으로 LSTM을 사용하여 단위 가격을 예측하였다. 입력 변수로는 농산물 거래 데이터, 기상 데이터, 수출입 데이터, 물가 상승지수 데이터를 사용해 14일 후 단위 가격을 예측하였으며 R2 Score 기준 마늘은 0.8798, 배추는 0.92935, 양파는 0.97388의 성능을 보였다. 이는 LSTM 기반의 농산물 가격 예측 모형은 높은 성능을 보이고 있으며, 다른 품종에 관해서도 연구해볼 가치가 있다는 것을 의미한다.

강민철과 이승호(2019)는 기후자료와 단위 면적당 감귤 생산량 자료를 이용하여 기후가 감귤 생산량에 미치는 영향을 파악하고자 하였다. 생육 시기별 기후요소와 단위 면적당 감귤 생산량 증감분 간의 관계를 분석하고, 기후변동이 생산량 변동에 미치는 영향을 파악하였다.

김세원과 김영희(2021)는 3가지의 머신러닝 알고리즘(Ridge Regression, Random Forest, XGBoost)을 후보 알고리즘으로 선정하여 작물 생산량 예측의 적합도를 평가 분석하고, 알고리즘별 최적의 모수(parameter)를 찾았다. 이는 예측 정확도를 우선시하는 심층학습 방법을 적용하지 않았다는 한계점을 가지고 있다.

정대호와 조영열(2022)은 딥러닝 기반의 LSTM 네트워크를 이용하여 9개의 도매시장에서 거래되고 있는 한라봉의 3kg 단위 평균 가격을 예측했다. 이는 9개

도매시장에서 거래되는 데이터만을 사용했다는 점과 다양한 감귤 품종 중 한라봉만 사용했다는 한계점을 가지고 있다.

박종현 등(2023)은 Multi-Step Time Series의 세 가지 전략을 비교 분석하기 위해 LGBM, MLP, LSTM, GRU를 사용하여 농산물(양파, 오이, 배추)의 중장기 가격 예측에 대한 최적의 모형을 제안하였다. 이 연구에서는 농산물의 가격 데이터만을 가지고 예측 모델을 만들었으므로 지역 현황, 시장 현황 등 외부 환경 요인을 전혀 고려하지 못했다는 한계점을 가지고 있다.

이에 본 연구에서는 작물 생산과 가격 예측 논문의 기법들을 참고하여 제주 감귤 맞춤형으로 개선하고, 제주 밀감의 대부분을 차지하는 노지온주와 하우스온주에 대해 외부 환경 요인을 고려한 출하량 및 가격 예측 모형을 개발하려고 한다.

1.2. 연구 목적

1.2.1. 연구 대상

제주 감귤에는 노지온주, 월동온주, 하우스온주, 한라봉, 천혜향, 황금향, 설국향, 레드향, 카레향 등 수많은 품종이 존재한다. 본 연구에서는 이 중에서 대표적인 노지온주와 하우스온주를 대상으로 출하량과 가격을 예측하는 모형을 개발할 것이다.

강민철과 이승호(2019)에 따르면 제주도의 기온 변동은 감귤 생산량 변동에 큰 영향을 미치지 못하지만, 일조 시간과 풍속, 강수량, 강수일수, 적설량 등 다른 기상 요인들의 중요성이 주목받았다. 또한 최근 심해진 기후변화는 감귤의 생산성에 직접적인 영향을 미칠 수 있어서 제주도에서 감귤 수급 안정성을 강화하기 위해 고려해야 할 중요 변수이다. 따라서 본 연구에서 개발하는 감귤 출하량 예측 모델은 시간에 따른 다양한 기상 요건들을 독립변수로 하여 출하량을 예측하는 모형으로 설계했다.

농산물 수입시장 개방과 더불어 과일 종류가 다양화되면서 제주 감귤에 대한 수요가 점차 낮아지는 추세이다. 이는 감귤 거래량과 가격 변화에 국내 타과일(경쟁 과일) 거래량과 해외 과일 수입량이 매우 큰 영향요인으로 작용하고 있다

는 것을 의미한다. 따라서 본 연구에서는 감귤 가격 변화에 영향을 미치는 경쟁 과일을 찾고, 감귤 가격 예측 모델은 해당 경쟁 과일의 수입량, 거래량, 거래 금액과 공휴일 정보 등을 시간에 따라 독립변수로 설정하여 감귤의 가격을 예측하는 모형으로 설계했다.

1.2.2. 연구 범위

본 연구는 노지온주와 하우스온주 감귤에 대해 심층학습 모델을 사용하여 출하량과 가격을 예측하고 윈도우 사이즈(window size)에 따른 적용 시나리오에 따라 모델 간의 성능을 비교할 것이다. 이때 사용되는 심층학습 모델로는 LSTM, Seq2Seq, Attention 기반 모델 세 종류를 사용하였다.

본 연구에서 활용할 데이터는 기상 데이터, 감귤 출하량 데이터, 과일 도매가격 및 거래량 데이터, 과일 수입/수출 데이터, 공휴일 데이터이다. 데이터마다 기준 일자 단위(일간, 월간 등)가 다르므로 모든 데이터는 주 단위로 변환하여 집계한 후 사용하게 되며 감귤의 출하량과 가격 예측 성능을 올리기 위해 다양한 파생 변수를 생성하여 출하량 및 가격 예측에 최적화된 데이터 마트를 구성하고 학습 데이터로 가공하여 모델을 학습시켰다.

또한, 각 데이터에 대해 탐색적 데이터 분석(Exploratory Data Analysis, EDA)을 통해 시간 흐름에 따른 추이 변화와 항목별 분포를 분석하여 유의미한 인사이트를 도출하고 감귤 출하량과의 상관성, 감귤 가격과의 상관성이 높은 특성을 찾아 예측 모델에 반영하였다.

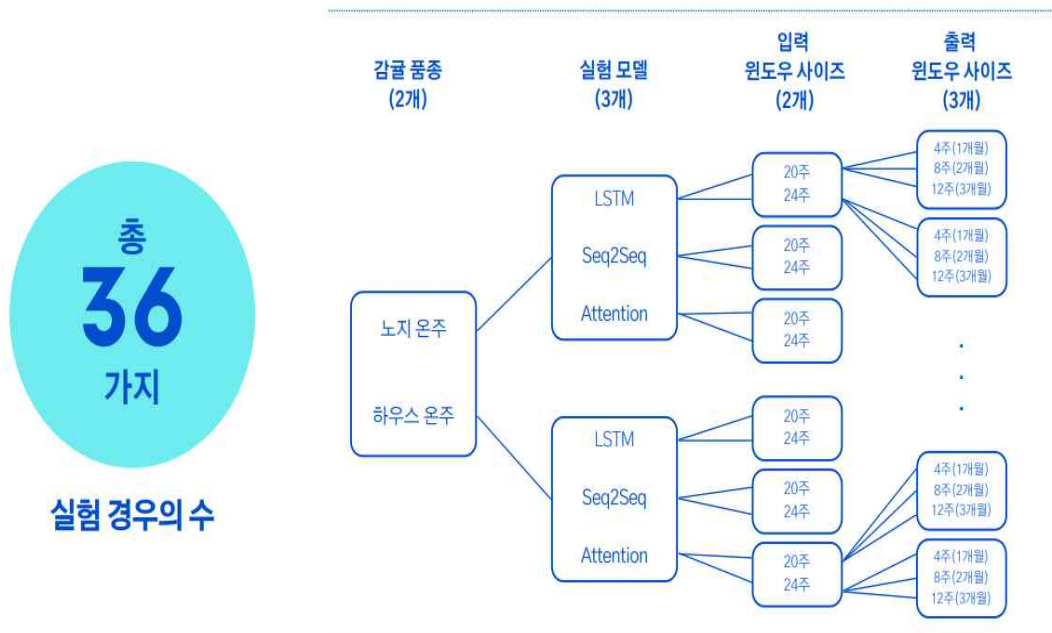
2. 연구 방법

2.1. 감귤 출하량 및 가격 예측 모형

2.1.1. 감귤 출하량 예측 모형

본 연구에서 개발한 출하량 예측 모형은 기상 데이터와 과거 출하량 데이터를 입력으로 받아 노지온주와 하우스온주의 출하량을 예측하는 모형이다.

[그림2]와 같이 출하량 예측 모형은 감귤 품종 2종, 시계열 모델 3개, 윈도우 사이즈는 입력 기간 20주(5개월), 24주(6개월), 출력 기간 4주(1개월), 8주(2개월), 12주(3개월)로 설정하여 총 36가지 경우에 대해 실험을 진행하였다.

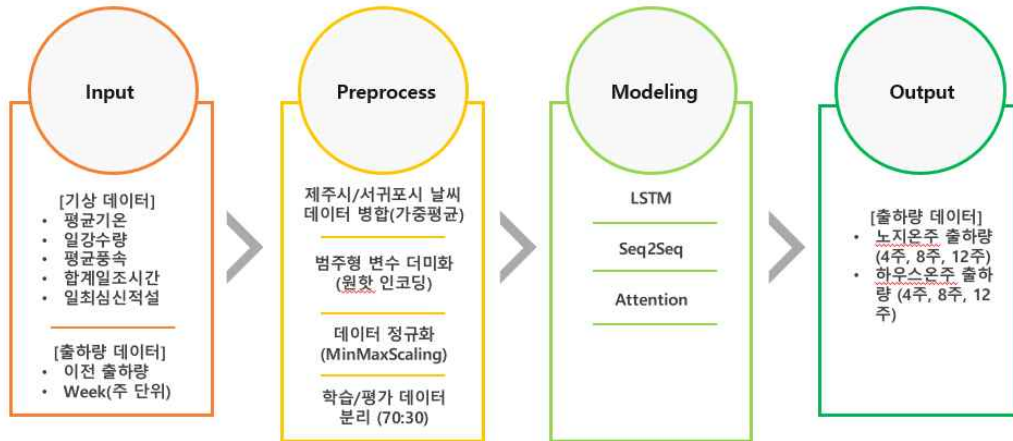


[그림 2] 감귤 출하량 예측 모형 실험 경우의 수

[그림3]과 같이 입력 데이터로는 기상 데이터의 날씨 항목들과 출하량 데이터의 과거 출하량을 주 단위로 집계하여 사용하였으며, 제주시와 서귀포시를 구분하여 날씨 항목에 대해 가중평균을 적용하고 범주형 변수 더미화, 데이터 정규화 등의 전처리를 진행하여 학습 데이터를 구성하였다. 출하량 예측 모형에는 LSTM, Seq2Seq, Attention 기반 모델을 모두 실험하고 윈도우 사이즈에 따른 모델 간의 성능을 비교하는 방법으로 실험을 진행하였다. 출하량 예측 모형의 구성

과 모델 개발에 사용된 주요 모수는 아래와 같다.

[모델 개발 프로세스]



[모델의 구성]

- 예측 모델: LSTM , Seq2Seq, Attention
- Input term: 20주, 24주 / Output term: 4주, 8주 ,12주
- 평가지표: RMSE, MAE, R_{adj}^2

[그림 3] 감귤 출하량 예측 모델 개발 프로세스 및 모델의 구성

- Activation Function(활성함수): ELU
- Optimizer(옵티마이저): Adam
- Epoch(에포크): 500
- Batch Size(배치 크기): 2
- Loss Function(손실함수): RMSE, MAE, R_{adj}^2 를 사용

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{식1})$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{식2})$$

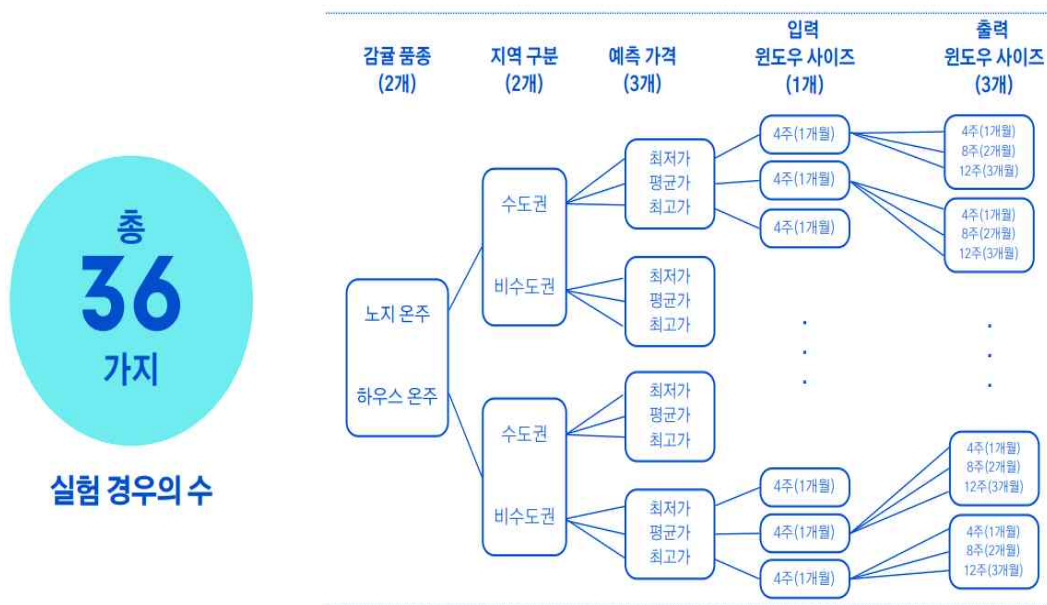
$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}, \quad R_{adj}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \quad (\text{식3})$$

(n: 데이터의 수, k: 독립변수의 수)

2.1.2. 감귤 가격 예측 모형

본 연구에서 개발한 가격 예측 모형은 감귤 출하량 데이터, 공휴일 데이터, 경쟁 과일의 가격 데이터, 감귤 가격 및 거래량 데이터를 입력으로 받아 노지온주와 하우스온주의 제주지역 최저가, 평균가, 최고가를 예측하는 모형이다.

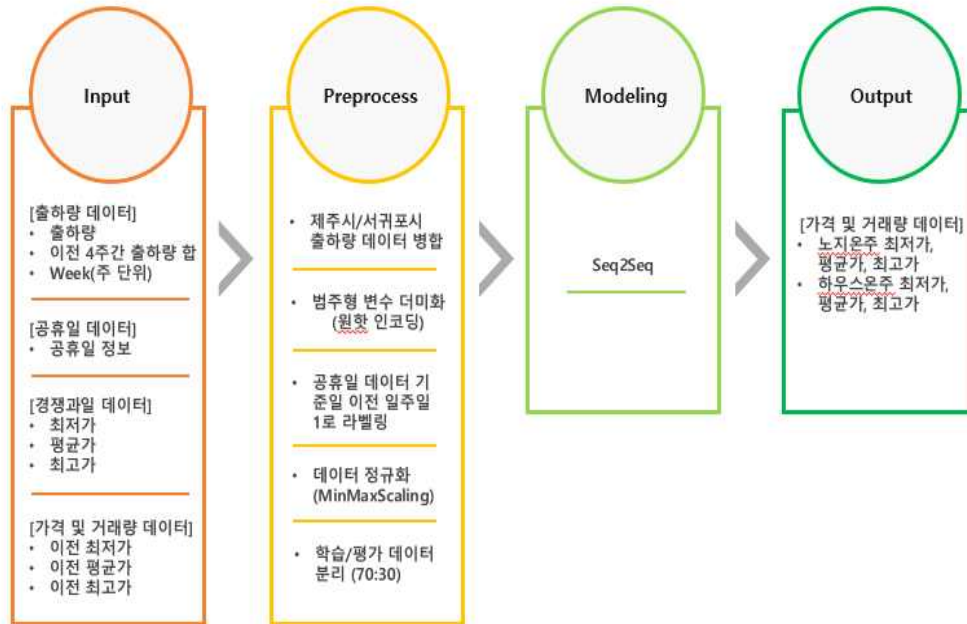
[그림4]와 같이 가격 예측 모형은 감귤 품종 2종, 지역 2지역(수도권, 비수도권), 예측 가격 3개(최저가, 평균가, 최고가), 시계열 모델 1개, 윈도우 사이즈는 입력 기간 4주(1개월), 출력 기간 4주(1개월), 8주(2개월), 12주(3개월)를 고려하여 총 36가지 경우에 대해 실험을 진행하였다.



[그림 4] 감귤 가격 예측 모형 실험 경우의 수

[그림5]와 같이 입력 데이터로는 출하량 데이터의 과거 출하량, 과일 가격 및 거래량 데이터에서 감귤과 경쟁 과일의 과거 최저가/평균가/최고가, 공휴일 데이터에서 가격에 영향을 미치는 공휴일 정보 등을 주 단위로 집계하여 사용하였으며, 출하량 데이터 병합, 범주형 변수 더미화, 공휴일 데이터 원핫 인코딩, 데이터 정규화 등의 전처리를 진행하여 학습 데이터를 구성하였다. 가격 예측 모형에는 Seq2Seq 모델을 사용하여 윈도우 사이즈에 따른 최저가, 평균가, 최고가의 예측 성능을 비교하는 방법으로 실험을 진행하였다. 가격 예측 모형의 구성은 아래와 같으며 모델 개발에 사용된 주요 모수는 출하량 모델과 동일하다.

[모델 개발 프로세스]



[모델의 구성]

- 예측 모델: Seq2Seq
- 지역 구분: 9대 도매시장 기준 수도권, 비수도권
- Input term: 4주 / Output term: 4주, 8주, 12주
- 예측 가격: 최저가, 평균가, 최고가
- 평가지표: RMSE, MAE, R_{adj}^2

[그림 5] 감귤 가격 예측 모델 개발 프로세스 및 모델의 구성

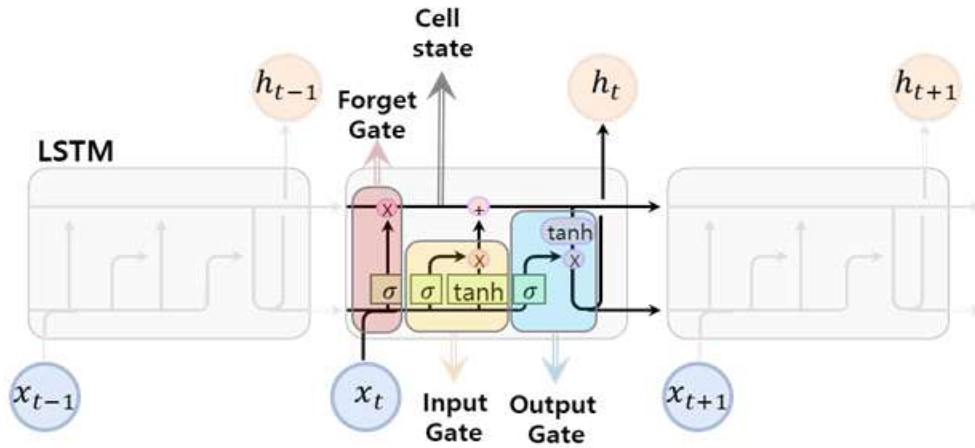
2.2. 심층학습 모형

2.2.1. LSTM 모델

Hochreiter와 Schmidhuber(1997)에 따르면 Long Short-Term Memory(LSTM) Network는 Recurrent Neural Network(RNN)의 입력되는 시퀀스의 길이가 길어질 수록 앞의 정보가 뒤로 충분히 전달되지 못하는 장기 의존성 문제를 해결하기 위해 고안된 모델이다.

이 모델은 [그림6]과 같이 시간에 따른 정보의 흐름을 새롭게 공급하는 Cell state를 정의하고 현재 시점의 정보를 바탕으로 과거 내용을 얼마나 잊고 기억할 지 등을 계산하여, 그 결과에 현재 정보를 추가해 다음 시점으로 정보를 전달하는 방식으로 이루어진다. 이때 LSTM은 3개의 Gate와 1개의 Cell state의 상호작용

용으로 이루어지며, 각 Gate는 정보를 저장하고 다음 단계로 전달하는 임무를 수행한다.



[그림 6] LSTM 모델 구조

- Forget gate: 과거의 정보를 얼마나 잊을지 혹은 기억할지 계산

$$f_t = \text{sigmoid}(W_f x_t + U_f h_{t-1} + b_f) \quad (\text{식4})$$

- Input gate: 현재 입력 정보를 얼마나 반영할지 계산

$$i_t = \text{sigmoid}(W_i x_t + U_i h_{t-1} + b_i) \quad (\text{식5})$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

- Output gate: 최종적으로 어떤 값을 출력할지 계산

$$o_t = \text{sigmoid}(W_o x_t + U_o h_{t-1} + b_o) \quad (\text{식6})$$

$$h_t = o_t \times \tanh(C_t)$$

- Cell state: 과거 Cell state를 새로운 state로 업데이트

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (\text{식7})$$

이 모델의 장점으로는 RNN의 장기 의존성 문제를 일부 해결하며, 과거의 정보를 훨씬 더 잘 반영한다. 또한, Cell state에 잊거나 기억해야 할 정보를 담아서 업데이트시키기 때문에 시퀀스가 길어져도 유지하려는 정보가 사라지지 않는다는 장점이 있으며, 단점은 모수가 많아 계산이 복잡하며 학습하는 데 시간이 오래 걸리고, 순서를 고려하기 쉽지 않은 문제점을 가지고 있다. (Hochreiter과 Schmidhuber, 1997; Olah, 2015)

본 연구에서 다루고 있는 출하량 및 가격 데이터는 시계열 데이터이므로 시계열 데이터의 예측에 있어 딥러닝의 Base model로 활용되는 LSTM을 선정하였다.

본 연구의 LSTM 기반 감귤 출하량 예측 모형의 레이어 구조는 [표5]와 같다.

[표 5] LSTM 기반 감귤 출하량 모형 레이어 구조

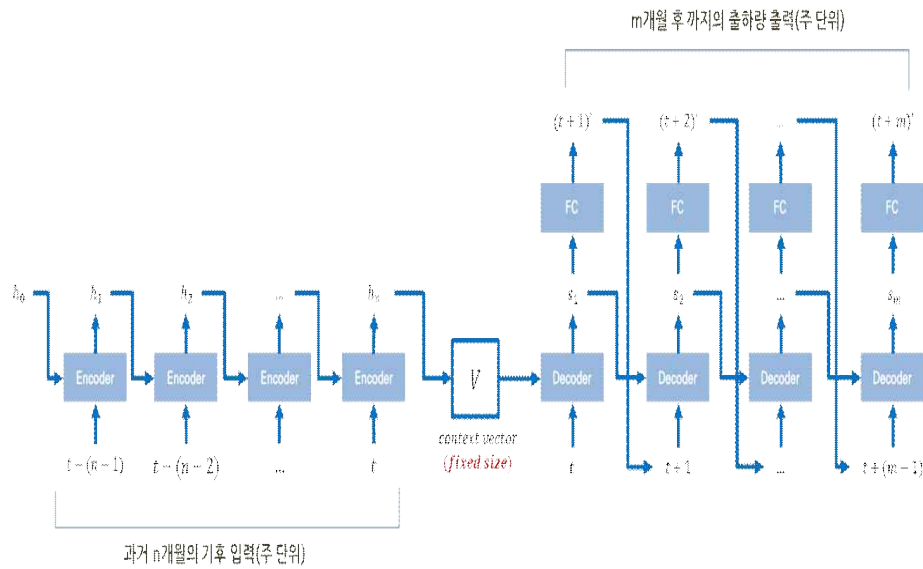
Layer	Input Shape	Parameters	Output Shape
0_lstm.lstm	(input_features=20or24, input_column=17)	(17, 512, batch_first=True)	(batch_size=2, input_features=20or24, hidden_state=512)
1_linear.Linear_0	-	Linear(in_features=512, out_features=12, bias=True)	(batch_size=2, out_features=4or8or12)
2_linear.ELU_1	-	(alpha=1.0)	(batch_size=2, out_features=4or8or12]
※ input_features와 output_features는 입출력 윈도우 사이즈에 따라 변동됨			

2.2.2. Seq2Seq 모델

Sutskever 등(2014)에 따르면 Sequence-to-Sequence(Seq2Seq) 모델은 연속형 데이터(시계열, 자연어 등)에 주로 사용되며, 입력된 시퀀스로부터 다른 도메인의 시퀀스를 출력하는 다양한 분야에서 사용되는 모델이다.

Seq2Seq 모델은 [그림7]과 같이 크게 인코더(Encoder)와 디코더(Decoder)라는 두 개의 모듈로 구성된다. 인코더는 입력으로 주어진 모든 x 에 대해 순차적으로 입력받은 뒤, 마지막에 이 모든 x 의 정보를 압축하여 하나의 컨텍스트 벡터(Context vector)를 생성하고, 디코더는 인코더로부터 컨텍스트 벡터를 받아 출력값을 순차적으로 출력하는데 학습과 평가 과정의 작동 방식은 아래와 같다.

- 학습(Train) 과정: 디코더에게 인코더에서 전달된 컨텍스트 벡터와 실제 값을 입력 받았을 때, 실제 값을 출력해야 한다고 학습함(Teacher forcing, 교사 강요)
- 평가(Test) 과정: 컨텍스트 벡터와 디코더의 입력값을 받아 다음에 올 시퀀스를 예측하고, 그 예측한 시퀀스를 다음 시점의 입력으로 넣는 행위를 반복



[그림 7] Seq2Seq 모델 구조

이 모델의 장점으로는 전체 입력을 살펴본 후, 임의의 컨텍스트 벡터를 출력함으로써 데이터에 대한 전체적인 맥락을 파악할 수 있으며, 단점으로는 하나의 고정된 크기의 벡터에 모든 정보를 압축함으로써 정보 손실이 발생하고 RNN의 고질적인 문제인 장기 의존성 문제가 여전히 존재한다. (Sutskever 등, 2014; Bai 등, 2018)

본 연구에서는 출하량 및 가격의 예측 정확도 향상을 위해 활용되었으며, 과거 n 개의 입력 데이터를 단일 값이 아닌 Sequence 형태로 받아 미래 m 개의 출력 데이터를 Sequence로 도출함으로써 기존 시계열 모형에서의 한계점을 보완하기 위해 해당 모델을 선정하였다.

본 연구의 Seq2Seq 기반 감귤 출하량 예측 모형의 레이어 구조는 [표6]과 같으며, 감귤 가격 예측 모형의 레이어 구조는 [표7]과 같다.

[표 6] Seq2Seq 기반 감귤 출하량 예측 모형 레이어 구조

Layer	Input Shape	Parameters	Output Shape
0_Encoder.rnn.lstm	(enc_input_features=20or24, input_column=17)	(17, 512, batch_first=True)	(batch_size=2, enc_input_features=20or24, hidden_state=512)
1_Decoder.rnn.lstm	-	(1, 512, batch_first=True)	(batch_size=2, enc_input_features=20or24, hidden_state=512]
2_Decoder.linear.Linear_0	-	(in_features=512, out_features=1, bias=True)	(batch_size=2, dec_output_features=1)
3_Decoder.linear.ELU_1	-	(alpha=1.0)	(batch_size=2, dec_output_features=1)
4_Decoder.linear.ELU_1	-	(alpha=1.0)	(batch_size=2, dec_output_features=1)

※ enc_input_features는 입력 윈도우 사이즈에 따라 변동됨

[표 7] Seq2Seq 기반 감귤 가격 예측 모형 레이어 구조

Layer	Input Shape	Parameters	Output Shape
0_Encoder.rnn.lstm	(enc_input_features=20or24, input_column=16or17or19or21)	(16or17or19or21, 512, batch_first=True)	(batch_size=2, enc_input_features=20or24, hidden_state=512)
1_Decoder.rnn.lstm	-	(1, 512, batch_first=True)	(batch_size=2, enc_input_features=20or24, hidden_state=512]
2_Decoder.linear.Linear_0	-	(in_features=512, out_features=1, bias=True)	(batch_size=2, dec_output_features=1)
3_Decoder.linear.ELU_1	-	(alpha=1.0)	(batch_size=2, dec_output_features=1)
4_Decoder.linear.ELU_1	-	(alpha=1.0)	(batch_size=2, dec_output_features=1)

※ enc_input_features, input_column은 입력 윈도우 사이즈와 지역(수도권/비수도권), 품종(노지온주/하우스온주)에 따른 경쟁 과일에 따라 변동됨

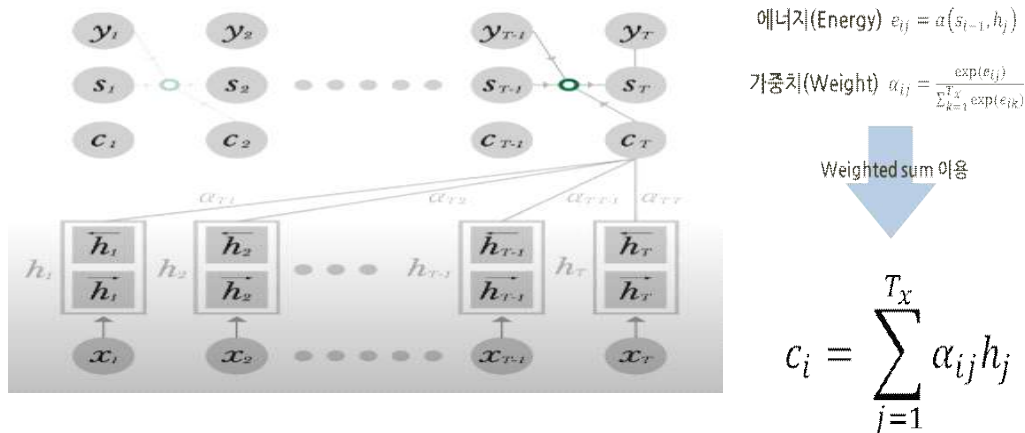
2.2.3. Attention 기반 모델

Vaswani 등(2017)에 따르면 Attention 모델의 메커니즘은 Seq2Seq 방법의 하나이며, Seq2Seq 방법론의 한계점인 여전히 RNN의 장기 의존성 문제가 모두 해결되지 않는 부분과 항상 고정된 크기의 벡터에 정보를 함축시킴으로써 정보 손실

이 되는 부분을 개선한 모델이다.

Attention 모델의 구조는 [그림8]과 같으며 Seq2Seq 모델과 마찬가지로 인코더와 디코더, 두 개의 모듈로 구성되어 있으나 중요한 부분에 더 집중(Attention)해서 학습한다는 점에서 Seq2Seq 모델과 다르게 작용한다.

- 인코더: 예측하려고 하는 값에 어떤 Sequence가 중요한지 가중치를 곱해서 새로운 input을 구성하고 컨텍스트 벡터 생성
- 디코더: 매번 Encoder의 모든 출력 중에서 어떤 정보가 중요한지를 계산. 어떤 시점을 많이 반영할 것인지 계산하여 출력
- 컨텍스트 벡터(c_i): 가중치를 실제 각 입력값과 hidden state와 곱해주고, 이를 모두 더해준 값을 디코더의 input으로 사용



[그림 8] Attention 모델 구조(左) 및 컨텍스트 벡터 정의(右)

이 모델의 장점으로서는 Seq2Seq의 한계점을 해결하여 과거에 예측하고자 하는 시점의 값과 비슷한 경험이 존재하는 경우 예측 성능이 뛰어나고 인코더와 디코더의 가중치를 시각화 및 확인할 수 있음으로써 기존 인공지능망의 설명력 부족 문제를 일부 해결할 수 있다.

이 모델의 단점은 시퀀스의 길이가 길어질수록, 입력을 순차적으로 처리하기 때문에 병렬 처리할 수 없음으로써 연산시간이 급속히 늘어나고, 이전에 고려하는 시점의 데이터가 현재 데이터 패턴과 맞지 않을 경우 학습이 잘되지 않는 경향이 있다. (Vaswani 등, 2017; Olah과 Carter, 2016; Qin 등, 2017)

본 연구의 Attention 기반 감귤 출하량 예측 모형의 레이어 구조는 [표8]과 같다.

[표 8] Attention 기반 감귤 출하량 예측 모형 레이어 구조

Layer	Input Shape	Parameters	Output Shape
0_Encoder.rnn.gru	(enc_input_features=20or24, input_column=17)	(17, 512)	(batch_size=2, enc_input_features=20or24, hidden_state=512)
1_Encoder.fc.Linear	-	(in_features=512, out_features=512, bias=True)	(batch_size=2, enc_input_features=20or24, hidden_state=512)
2_Encoder.act.ELU_1	-	(alpha=1.0)	(batch_size=2, enc_input_features=20or24, hidden_state=512)
1_Decoder.attention.attn.Linear_0	-	(in_features=1024, out_features=512, bias=True)	(batch_size=2, enc_output_features=1, hidden_state=512)
2_Decoder.attention.v.Linear_0	-	(in_features=512, out_features=1, bias=False)	(batch_size=2, dec_output_features=1)
3_Decpder.rnn.gru	-	(513, 512)	(batch_size=2, hidden_state=512)
4_Decpder.fc_out.Linear_0	-	(in_features=1025, out_features=1, bias=True)	(batch_size=2, dec_output_features=1)
※ enc_input_features는 입력 윈도우 사이즈에 따라 변동됨			

3. 연구 자료

3.1. 활용 데이터

3.1.1. 기상 데이터

기상청에서 설치한 종관기상 관측 장비에 의해 수집된 일별 기후 정보 데이터로 공공데이터포털의 기상청 API를 활용하여 데이터를 수집했다. (기상청, 2021) 2017년 1월부터 2022년 12월까지 약 6년간의 제주지역의 기후 데이터 4,382건을 수집하여 데이터 마트를 구축하였고 이 데이터는 [표9]와 같이 총 68개의 변수로 구성되어 있었다.

[표 9] 기상청 API 호출 결과

항목명(영문)	항목명(한글)	설명
numOfRows	한 페이지 결과 수	한 페이지당 표출 데이터 수
pageNo	페이지 번호	페이지 수
totalCount	데이터 총 개수	데이터 총 개수
resultCode	응답메시지 코드	응답 메시지코드
resultMsg	응답메시지 내용	응답 메시지 설명
dataType	데이터 타입	응답자료형식 (XML/JSON)
stnId	지점 번호	종관기상관측 지점 번호
stnNm	지점명	종관기상관측 지점명
tm	시간	일시
avgTa	평균 기온	평균 기온(°C)
minTa	최저 기온	최저 기온(°C)
minTaHrmt	최저 기온 시각	최저 기온 시각(hhmi)
maxTa	최고 기온	최고 기온(°C)
maxTaHrmt	최고 기온 시각	최고 기온 시각(hhmi)
sumRnDur	강수 계속시간	강수 계속시간(hr)
mi10MaxRn	10분 최다강수량	10분 최다강수량(mm)
mi10MaxRnHrmt	10분 최다강수량 시각	10분 최다강수량 시각(hhmi)
hr1MaxRn	1시간 최다강수량	1시간 최다강수량(mm)
hr1MaxRnHrmt	1시간 최다 강수량 시각	1시간 최다 강수량 시각(hhmi)
sumRn	일강수량	일강수량(mm)
maxInsWs	최대 순간풍속	최대 순간풍속(m/s)
maxInsWsWd	최대 순간 풍속 풍향	최대 순간 풍속 풍향(16방위)
maxInsWsHrmt	최대 순간풍속 시각	최대 순간풍속 시각(hhmi)
maxWs	최대 풍속	최대 풍속(m/s)
maxWsWd	최대 풍속 풍향	최대 풍속 풍향(16방위)
maxWsHrmt	최대 풍속 시각	최대 풍속 시각(hhmi)
avgWs	평균 풍속	평균 풍속(m/s)
hr24SumRws	풍정합	풍정합(100m)
maxWd	최대 풍향	최대 풍향(16방위)
avgTd	평균 이슬점온도	평균 이슬점온도(°C)
minRhm	최소 상대습도	최소 상대습도(%)
minRhmHrmt	평균 상대습도 시각	평균 상대습도 시각(hhmi)
avgRhm	평균 상대습도	평균 상대습도(%)
avgPv	평균 증기압	평균 증기압(hPa)
avgPa	평균 현지기압	평균 현지기압(hPa)
maxPs	최고 해면 기압	최고 해면 기압(hPa)
maxPsHrmt	최고 해면기압 시각	최고 해면기압 시각(hhmi)
minPs	최저 해면기압	최저 해면기압(hPa)
minPsHrmt	최저 해면기압 시각	최저 해면기압 시각(hhmi)
avgPs	평균 해면기압	평균 해면기압(hPa)
ssDur	가조시간	가조시간(hr)
sumSsHr	합계 일조 시간	합계 일조 시간(hr)
hr1MaxlcsrHrmt	1시간 최다 일사 시각	1시간 최다 일사 시각(hhmi)
hr1Maxlcsr	1시간 최다 일사량	1시간 최다 일사량(MJ/m2)
sumGsr	합계 일사량	합계 일사량(MJ/m2)
ddMefs	일 최심신적설	일 최심신적설(cm)

ddMefsHrmt	일 최심신적설 시각	일 최심신적설 시각(hhmi)
ddMes	일 최심적설	일 최심적설(cm)
ddMesHrmt	일 최심적설 시각	일 최심적설 시각(hhmi)
sumDpthFhsc	합계 3시간 신적설	합계 3시간 신적설(cm)
avgTca	평균 전운량	평균 전운량(10분위)
avgLmac	평균 중하층운량	평균 중하층운량(10분단위)
avgTs	평균 지면온도	평균 지면온도(°C)
minTg	최저 초상온도	최저 초상온도(°C)
avgCm5Te	평균 5cm 지중온도	평균 5cm 지중온도(°C)
avgCm10Te	평균 10cm 지중온도	평균 10cm 지중온도(°C)
avgCm20Te	평균 20cm 지중온도	평균 20cm 지중온도(°C)
avgCm30Te	평균 30cm 지중온도	평균 30cm 지중온도(°C)
avgM05Te	0.5m 지중온도	0.5m 지중온도(°C)
avgM10Te	1.0m 지중온도	1.0m 지중온도(°C)
avgM15Te	1.5m 지중온도	1.5m 지중온도(°C)
avgM30Te	3.0m 지중온도	3.0m 지중온도(°C)
avgM50Te	5.0m 지중온도	5.0m 지중온도(°C)
sumLrgEv	합계 대형증발량	합계 대형증발량(mm)
sumSmlEv	합계 소형증발량	합계 소형증발량(mm)
n99Rn	9-9강수	당일 오전 9시부터 다음날 오전 9시까지 강수량(mm)
iscs	일기현상	일기현상
sumFogDur	안개 계속 시간	안개 계속 시간(hr)

이 데이터 중 동일한 값이 기록되어 있는 변수, 빈 값을 다수 포함하고 있는 변수 등을 제거하여 [표10]과 같이 주요 변수 14개를 선정하였다.

[표 10] 기상 정보 주요 변수 목록

번호	항목명(영문)	항목명(한글)
1	BASE_DATE	기준날짜
2	TEMPERATURE_MEAN	평균기온
3	TEMPERATURE_MIN	최저기온
4	TEMPERATURE_MAX	최고기온
5	RAINING_AMOUNT_DATE	일강수량
6	WIND_SPEED_INSTANT_MAX	최대순간풍속
7	WIND_SPEED_MEAN	평균풍속
8	HUMIDITY_MIN	최소상대습도
9	HUMIDITY_MEAN	평균상대습도
10	SUNSHINE_DURATION_POSSIBLE	가조시간
11	SUNSHINE_DURATION_SUM	합계일조시간
12	SNOW_AMOUNT_NEW	일최심신적설
13	SNOW_AMOUNT	일최심적설
14	GROUND_TEMPERATRUE_MEAN	평균지면온도

3.1.2. 감귤 출하량 데이터

감귤출하신고시스템을 통해 수집한 일별 감귤 출하량 정보 데이터로 제주감귤출하연합회에서 데이터를 수집했다. (제주감귤출하연합회, 2022) 2017년 1월부터 2022년 12월까지 약 6년간의 출하량 데이터 1,427,332건을 수집하여 데이터 마트를 구축하였고 이 데이터는 [표11]과 같이 총 9개의 변수로 구성되어 있었다.

[표 11] 감귤 출하량 데이터 변수 목록

번호	항목명(영문)	항목명(한글)
1	BASE_DATE	기준날짜
2	SHIPMENT_AREA_CODE	출하지역코드
3	SHIPMENT_GROUP_CODE	출하단체코드
4	SHIPMENT_REGISTER_PLACE_CODE	출하신고장소코드
5	CORPORATE_CODE	도매시장법인코드
6	KIND_CATEGORY	품종 카테고리
7	SHIPPING_STANDARD_CATEGORY	출하규격
8	BOX_NUMBER	상자수
9	TRADE_AMOUNT	출하량

3.1.3. 과일 도매가격 및 거래량 데이터

도매유통정보시스템에 의해 수집된 일별 과일 품목별 도매가격 및 거래량 정보 데이터로 aT도매시장통합홈페이지에서 데이터를 수집했다. (한국농수산식품유통공사, 2023) 2017년 1월부터 2022년 12월까지 약 6년간의 감귤과 경쟁 과일의 도매가격 및 거래량 데이터 334,305건을 수집하여 데이터 마트를 구축하였고 이 데이터는 [표12]와 같이 총 8개의 변수로 구성되어 있었다.

[표 12] 과일 도매가격 및 거래량 데이터 변수 목록

번호	항목명(영문)	항목명(한글)
1	BASE_DATE	기준날짜
2	MARKET_CODE	도매시장별 코드
3	FRUIT_CODE	과일별 코드
4	TRADE_UNIT	거래단위
5	TRADE_AMOUNT	거래량
6	PRICE_MAX	최고가
7	PRICE_MIN	최저가
8	PRICE_MEAN	평균가격

3.1.4. 과일 수입/수출 데이터

관세청 수출입통계정보에서 수집된 월별 과일 품목별 수입/수출 중량 및 금액 정보 데이터로 공공데이터포털의 관세청 API를 활용하여 데이터를 수집했다. (관세청, 2022) 2017년 1월부터 2022년 11월까지 약 5년 11개월간의 경쟁 과일의 수입/수출 데이터 426건을 수집하여 데이터 마트를 구축하였고 이 데이터는 [표 13]과 같이 총 6개의 변수로 구성되어 있었다.

[표 13] 과일 수입/수출 데이터 변수 목록

번호	항목명(영문)	항목명(한글)
1	BASE_DATE	기준날짜
2	EXPORT_AMOUNT	수출금액
3	IMPOT_AMOUNT	수입금액
4	TRADE_PAYMENT	무역수지
5	EXPORT_WEIGHT	수출중량
6	CODE_HS	품목코드

이때 수집한 경쟁 과일은 수입 과일 중 감귤 판매에 영향력이 높은 경쟁 과일로 [표14]와 같이 6개의 과일을 선정했다.

[표 14] 수입 과일 중 경쟁 과일 목록

번호	CODE_HS(품목 코드)	과일명
1	806100000	포도(신선)
2	803900000	바나나(기타)
3	804300000	파인애플(신선,건조)
4	810500000	키위푸르트(신선)
5	805100000	오렌지(신선,건조)
6	804502000	망고(신선,건조)

3.1.5. 공휴일 데이터

천문우주정보에서 수집된 특일(국경일, 공휴일, 기념일, 24절기, 잡절) 정보 데이터로 공공데이터포털의 한국천문연구원 API를 활용하여 데이터를 수집했다. (한국천문원, 2022) 2017년 1월부터 2022년 12월까지 약 6년간의 특일 데이터 108건을 수집하여 데이터 마트를 구축하였고 이 데이터는 [표15]와 같이 총 10개의 변수로 구성되어 있었다.

[표 15] 한국천문연구원 API 호출 결과

항목명(영문)	항목명(한글)	설명
resultCode	결과코드	결과코드
resultMsg	결과메시지	결과메시지
numOfRows	한 페이지 결과 수	한 페이지 결과 수
pageNo	페이지 번호	페이지번호
totalCount	전체 결과 수	전체 결과 수
locdate	날짜	날짜
seq	순번	순번
dateKind	종류	특일정보의 분류(02: 기념일)
isHoliday	공공기관 휴일여부	공공기관 휴일여부
dateName	명칭	명칭

이 데이터 중 불필요한 변수를 제거하여 [표16]과 같이 주요 변수 4개를 선정하였다.

[표 16] 공휴일 데이터 변수 목록

번호	항목명(영문)	항목명(한글)
1	HOLIDAY_KIND	휴일종류
2	HOLIDAY_NAME	휴일명칭
3	PUBLIC_INSTITUTION_HOLIDAY_FLAG	공공기관 휴일여부
4	BASE_DATE	기준날짜

3.2. 데이터 가공 방법

3.2.1. 기상 데이터

본 연구에서 기상 데이터는 제주지역을 대상으로 수집했으나, 제주지역 내에서도 지역별로 제주시와 서귀포시에 따라 기상(기온, 강수량, 풍속, 습도 등) 차이가 커 모든 지역에 대해 제주시와 서귀포시로 그룹화하고, 지역별(제주시, 서귀포시) 평균 값이 결측인 항목의 경우 제거하였다.

- 제거 항목: 1.0m지중온도, 1.5m지중온도, 3.0m지중온도, 안개지속시간

이번 연구에서 진행하는 출하량과 가격 예측 모델은 주 단위 예측을 기준으로 하고 있으므로 데이터의 시간 단위를 일에서 주로 변환하였다.

3.2.2. 감귤 출하량 데이터

감귤 출하량 데이터에서 품종 카테고리 코드를 [표17]과 같이 과일명으로 라벨

링하고, 감귤 출하량 데이터의 품종 카테고리 코드에는 노지온주가 따로 구분되어 있지 않아 극조생, 조생온주, 보통에 해당하는 코드를 모두 노지온주로 취급하여 라벨링 하였다.

본 연구에서 사용되는 감귤 품종인 노지온주와 하우스온주 이외의 데이터는 모두 제거하여 노지온주과 하우스온주에 대한 출하량 데이터는 총 886,197건이 추출되었다.

[표 17] 감귤 품종 카테고리 코드와 과일명 매칭 정보

KIND_CATEGORY(품종 카테고리)	과일명
1101	극조생(노지온주)
1201	조생온주(노지온주)
1301	보통(노지온주)
1202	월동온주
1401	한라봉
1404	금감
1402	청견
1406	진지향
1408	천혜향
1403	네블오렌지
1411	황금향
1412	레드향
1413	한라향(세토미)
1414	남진해(카라향 굴로향)
1415	청희오렌지
1416	기타1
1417	진지휘(탐라향)
1405	이예감
1407	하귤
1409	세미놀
1410	하루미
1400	기타
2201	하우스온주
1418	윈터프린스
1419	아스미

또한, 추출된 데이터에서 일 단위 데이터를 주 단위로 변환하고 기준날짜를 기준으로 연도와 월 정보를 파생변수로 생성하였다.

마지막으로 기상 데이터와 마찬가지로 지역을 제주시와 서귀포시로 구분하기 위해 지역코드가 '제주'와 '제주(구, 북군)'를 제주시로, '서귀포'와 '서귀포(구, 남

군)'를 서귀포시로 그룹화하는 작업을 진행하였다.

3.2.3. 과일 도매가격 및 거래량 데이터

감귤 도매가격 및 거래량 데이터에서 과일별 코드를 [표18]과 같이 한글명으로 라벨링하는 작업을 진행하였다.

[표 18] 과일별 코드와 과일명 매칭 정보

FRUIT_CODE(과일별 코드)	과일명
1	사과
2	배
3	단감
4	수입오렌지
5	딸기
6	수박
7	참외
8	포도
9	노지온주
10	하우스온주
11	월동온주
12	한라봉
17	천혜향
27	황금향
28	레드향
30	카라향

본 연구에서는 수도권과 비수도권을 나눠 감귤 가격을 예측하기 위해 제주감귤출하연합회에서 제공하는 31개 도매시장을 [표19]와 같이 지역별 9대 도매시장으로 구분 후 수도권과 비수도권에 따라 나눠서 라벨링 작업을 진행하였다. (제주감귤출하연합회, 2022)

[표 19] 도매시장별 구분 정보

MARKET_CODE (도매시장별 코드)	지역도매시장	수도권/ 비수도권	데이터 수
농협가락(공)	서울가락도매시장	수도권	22874
구리청과(주)	경기구리도매시장	수도권	13616
대구중앙청과	대구북부도매시장	비수도권	13594
중앙청과(주)	서울가락도매시장	수도권	13239
농협구리(공)	경기구리도매시장	수도권	12995
농협부산(공)	부산엄궁도매시장	비수도권	12001
농협강서(공)	서울강서도매시장	수도권	11943
농협북대구(공)	대구북부도매시장	비수도권	11529
서울청과(주)	서울가락도매시장	수도권	11442
효성청과(주)	대구북부도매시장	비수도권	10927
농협대전(공)	대전오정도매시장	비수도권	10773
광주원협(공)	광주각화도매시장	비수도권	10736
서부청과(주)	서울강서도매시장	수도권	10456
부산청과(공)	부산엄궁도매시장	비수도권	10440
광주청과(주)	광주각화도매시장	비수도권	10268
대양청과(주)	대구북부도매시장	비수도권	10212
덕풍청과(주)	인천남촌도매시장	수도권	9434
항도청과(주)	부산엄궁도매시장	비수도권	9084
동화청과(주)	서울가락도매시장	수도권	8633
경기청과(주)	경기수원도매시장	수도권	8589
대전청과(주)	대전오정도매시장	비수도권	8540
원협구월(공)	인천남촌도매시장	수도권	7611
대인농산(주)	인천남촌도매시장	수도권	7593
인천농산물(주)	인천남촌도매시장	수도권	7565
수원청과(주)	경기수원도매시장	수도권	7460
한국청과(주)	서울가락도매시장	수도권	7325
광주중앙청과	광주각화도매시장	비수도권	7281
경북원협(공)	대구북부도매시장	비수도권	7187
인터넷청과(주)	경기구리도매시장	수도권	6904
강서청과(주)	서울강서도매시장	수도권	5663
수원원협(공)	경기수원도매시장	수도권	4975

본 연구에서는 노지온주 5kg, 하우스온주 3kg 단위에 대해서만 가격 예측을 진행하므로 해당 거래단위 외의 데이터를 모두 제거하고, 거래단위와 거래량을 곱해 총 거래량 파생변수를 생성 후 데이터 기간을 일 단위에서 주 단위로 집계 하였다.

- TRADE_TOTAL(총 거래량, KG): 거래단위 * 거래량
- 거래 가격은 최저가는 주(7일)의 최저가, 최고가는 주(7일)의 최고가, 평균 가격은 주(7일)의 평균 가격으로 다운 샘플링하여 변환

- 총 거래량은 단순합을 통해 주 단위 변환
 - 가격에 대한 결측치가 발생할 경우, 결측치 '0'으로 대체
- 위에서 집계된 데이터에서 최저가, 최고가, 평균가가 평균 2σ 보다 크고 -2σ 보다 작은 데이터(95% 신뢰구간 적용)를 이상치로 보고 Nan으로 처리 후 선형보간법으로 Imputation 처리를 진행하였다.

3.2.4. 과일 수입/수출 데이터

과일 수입/수출 데이터에서 기준날짜를 기반으로 연도와 월 정보를 추출하여 파생변수를 생성하였다.

3.2.5. 공휴일 데이터

공휴일 데이터에서 모든 공휴일을 사용하지 않고, 감귤 거래에 영향을 미칠 것으로 예상되는 휴일만을 남기고 나머지 휴일은 제거하였다. 또한, 감귤 가격에 영향을 미치는 일자는 휴일 당일이 아닌 휴일 앞으로 7일간으로 가정하여 원핫 인코딩 처리를 통해 휴일 전 일주일의 데이터를 반영하였다.

- 추출된 휴일: 신정, 설날, 어린이날, 추석, 크리스마스

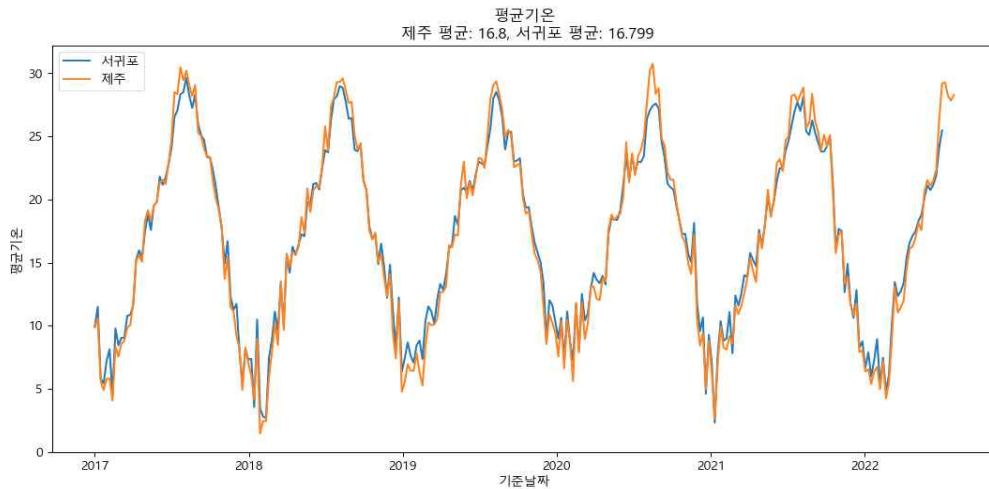
4. 분석 결과

4.1. 탐색적 데이터 분석

4.1.1. 기상 데이터 EDA

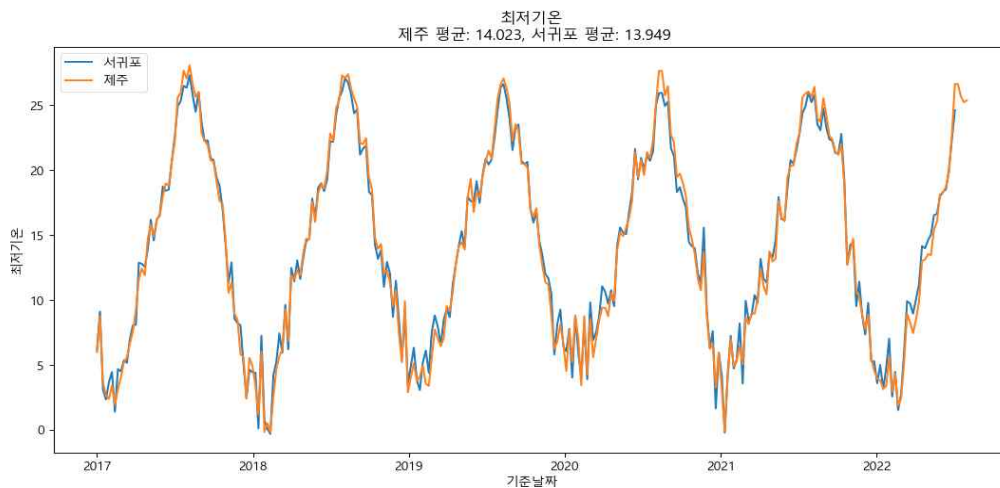
지역을 제주시와 서귀포시로 구분하여 주 단위 평균기온, 최저기온, 최고기온, 일강수량, 평균풍속, 평균상대습도, 일최심신적설, 일최심적성, 가조시간, 합계일조시간, 평균지면온도 변화 추이를 시각화하면 다음과 같은 기상 특징을 보인다.

- 평균기온: 제주시와 서귀포시가 거의 비슷함



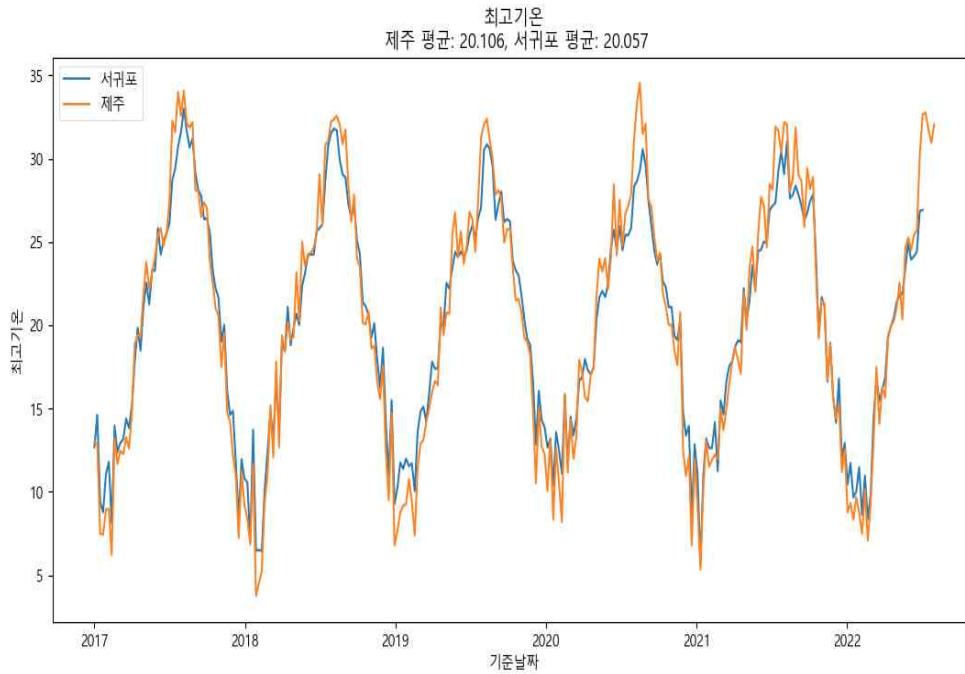
[그림 9] 지역별 평균기온(°C) 추이 시각화 결과

- 최저기온: 서귀포시가 제주시보다 미세하게 낮음



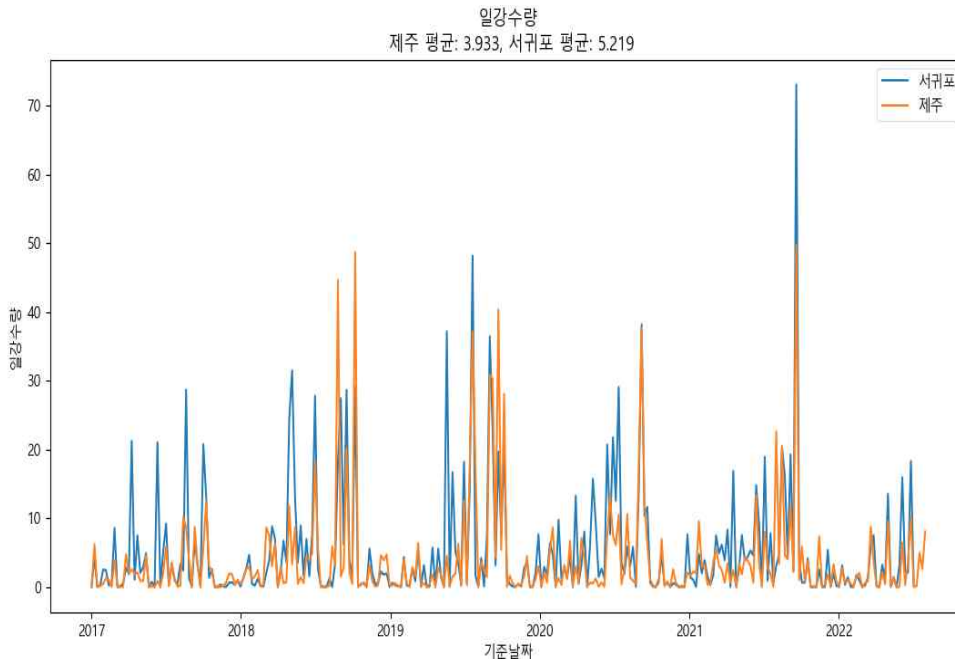
[그림 10] 지역별 최저기온(°C) 추이 시각화 결과

- 최고기온: 제주시가 서귀포시보다 미세하게 높음



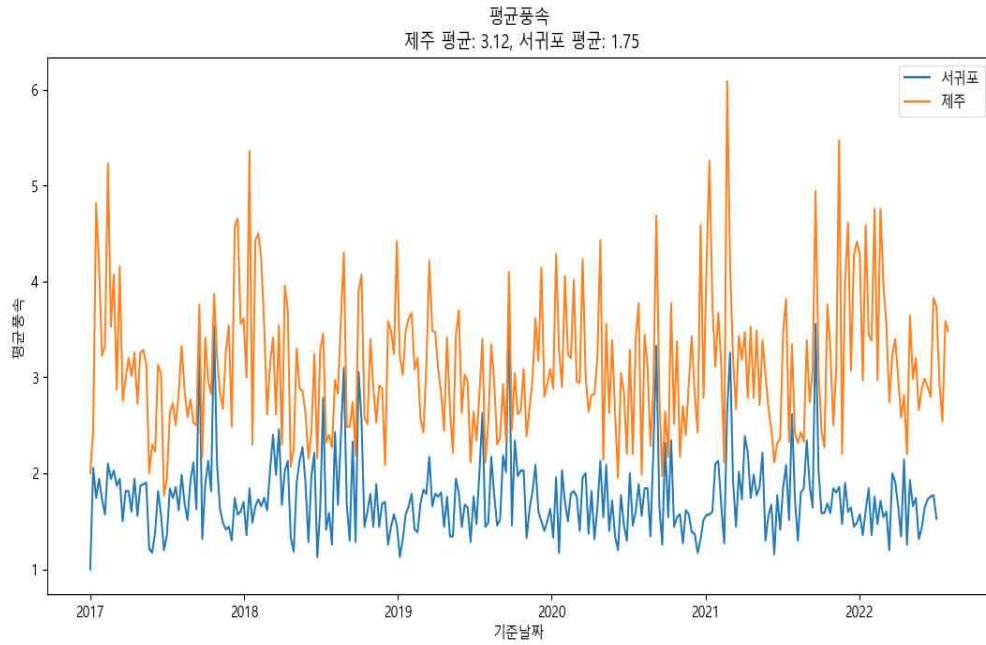
[그림 11] 지역별 최고기온(°C) 추이 시각화 결과

- 일강수량: 서귀포시가 제주시에 비해 많음



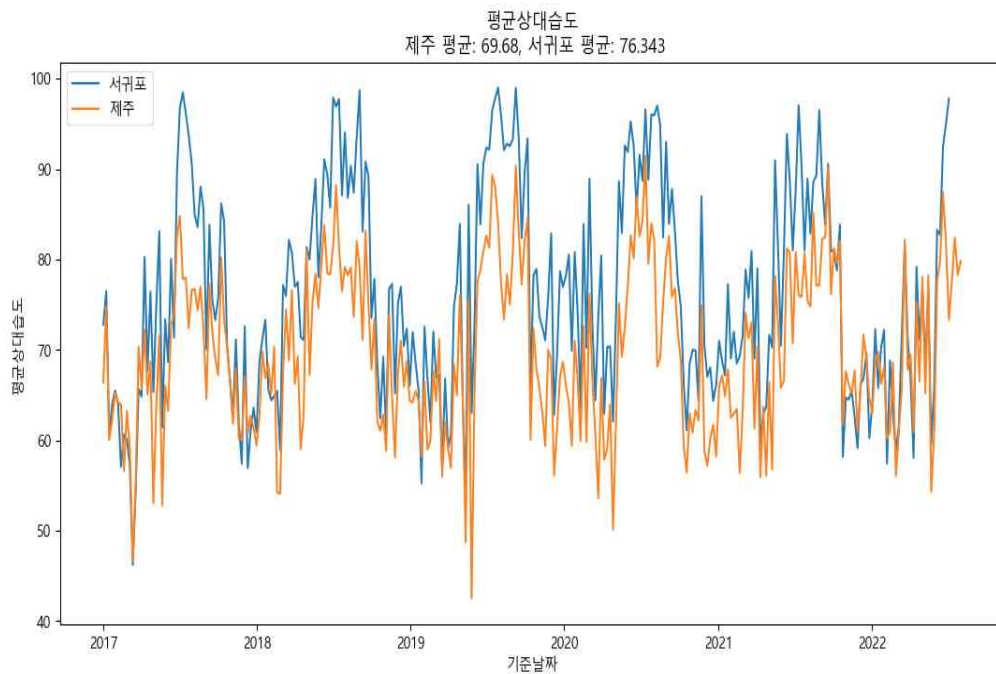
[그림 12] 지역별 일강수량(mm) 추이 시각화 결과

- 평균풍속: 제주시가 서귀포시에 비해 1.8배가량 빠름



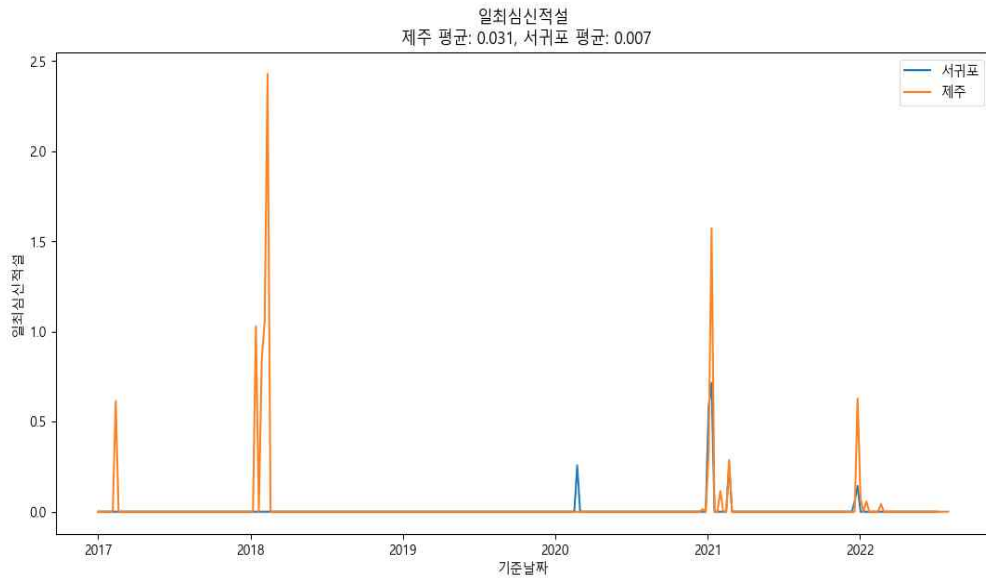
[그림 13] 지역별 평균풍속(m/s) 추이 시각화 결과

- 평균상대습도: 일강수량과 같이 서귀포시가 제주보다 높음



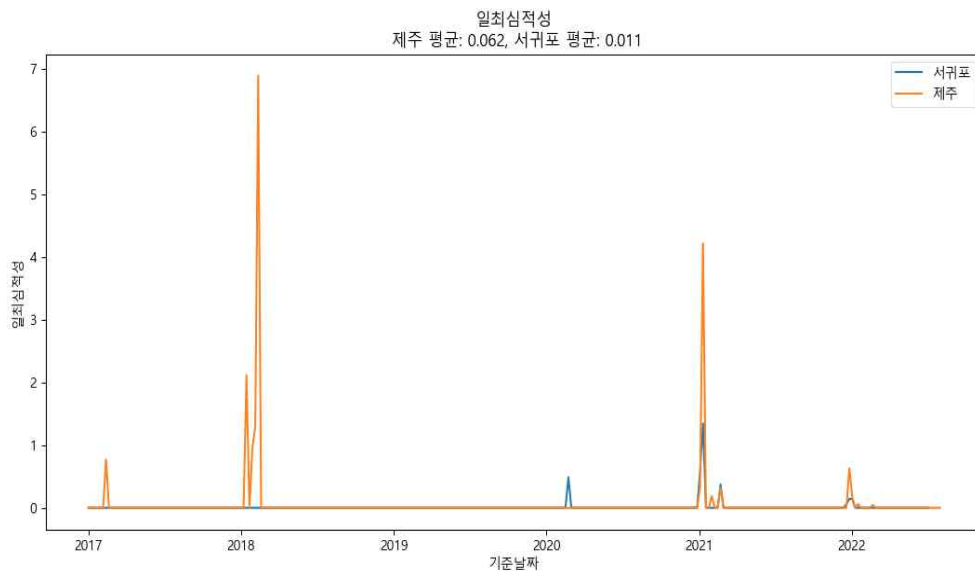
[그림 14] 지역별 평균상대습도(%) 추이 시각화 결과

- 일최심신적설: 일최심신적설은 기존에 내렸던 눈은 제거하고, 일정 기간 새롭게 쌓인 눈이 가장 두껍게 쌓여 있을 때 눈의 깊이를 나타내며 제주시가 서귀포시보다 평균 4.4배가량 높게 나타남. 서귀포시는 거의 눈이 오지 않음.



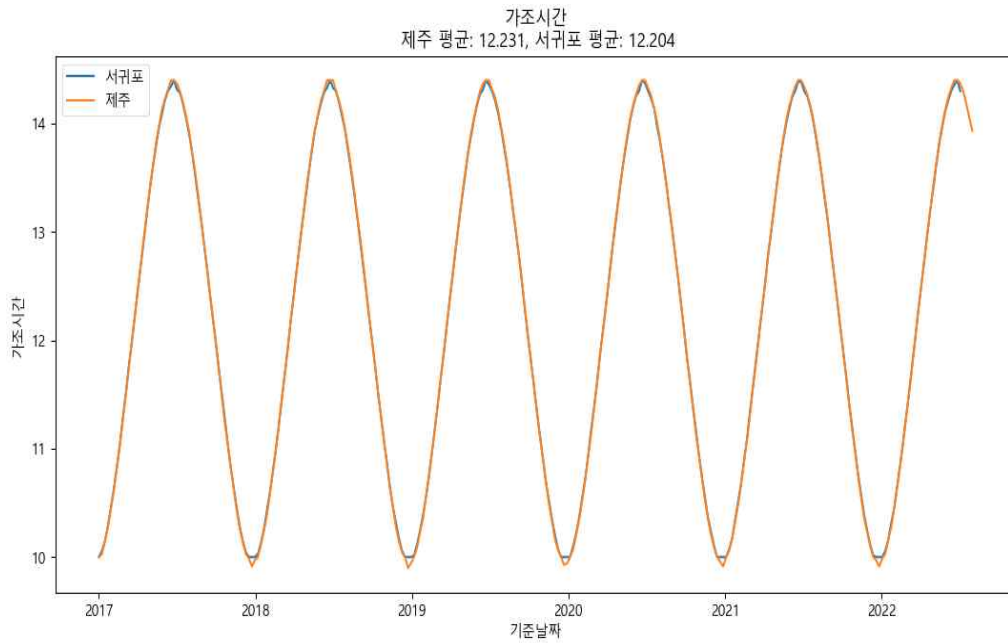
[그림 15] 지역별 일최심신적설(cm) 추이 시각화 결과

- 일최심적설: 일최심적설은 일정 기간에 전부터 내려서 녹지 않고 쌓여 있을 눈을 포함하여 가장 두껍게 쌓여 있을 때 눈의 깊이를 나타내며 제주시가 서귀포시보다 5.6배가량 높게 나타남. 서귀포시는 거의 눈이 오지 않음



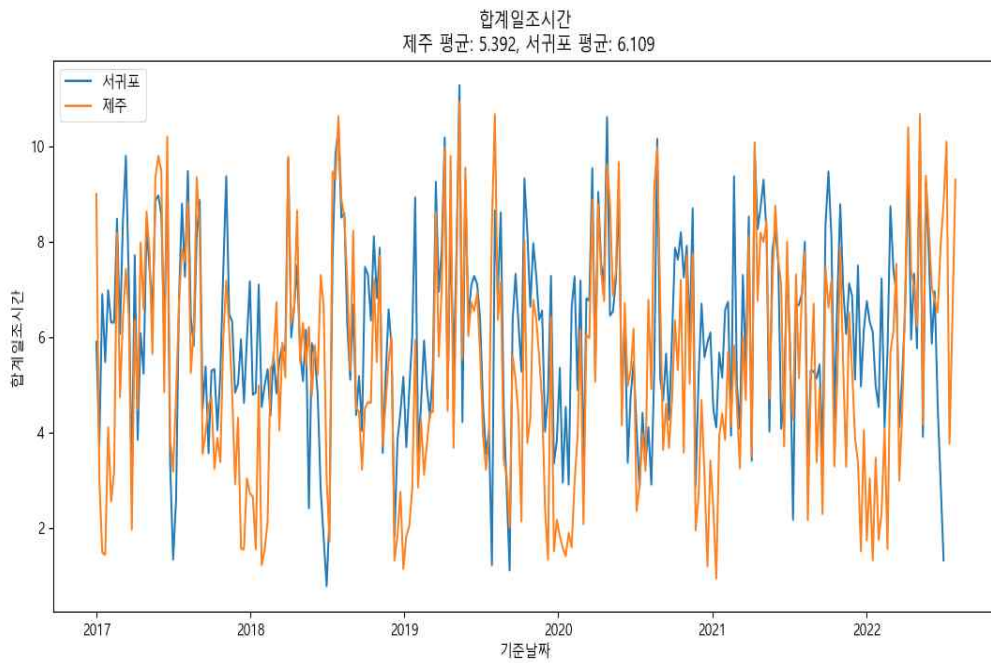
[그림 16] 지역별 일최심적설(cm) 추이 시각화 결과

- 가조시간: 제주와 서귀포시가 거의 비슷함



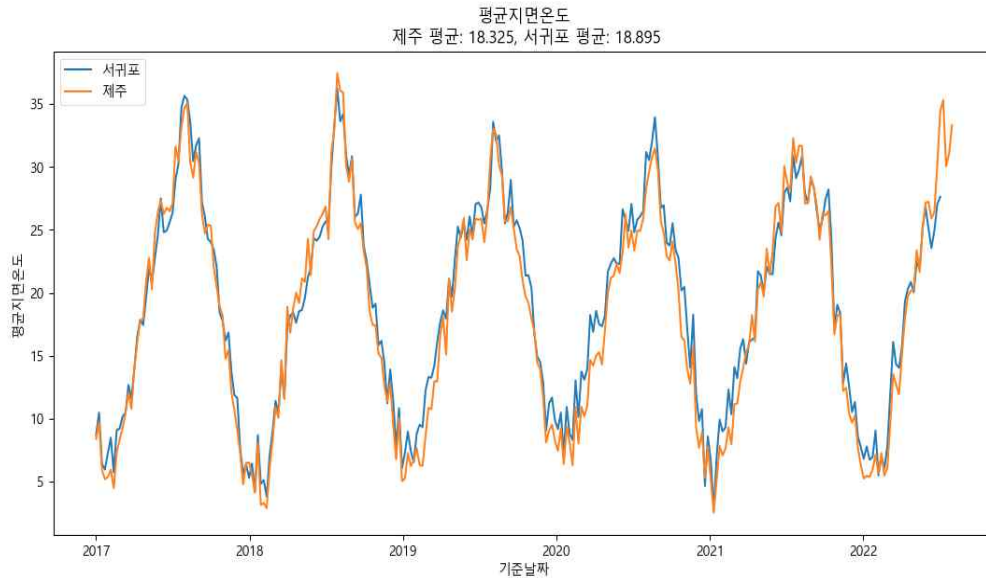
[그림 17] 지역별 가조시간(hr) 추이 시각화 결과

- 합계일조시간: 서귀포시가 제주시에 비해 좀 더 긴 편으로 나타남



[그림 18] 지역별 합계일조시간(hr) 추이 시각화 결과

- 평균지면온도: 제주시와 서귀포시가 거의 비슷함

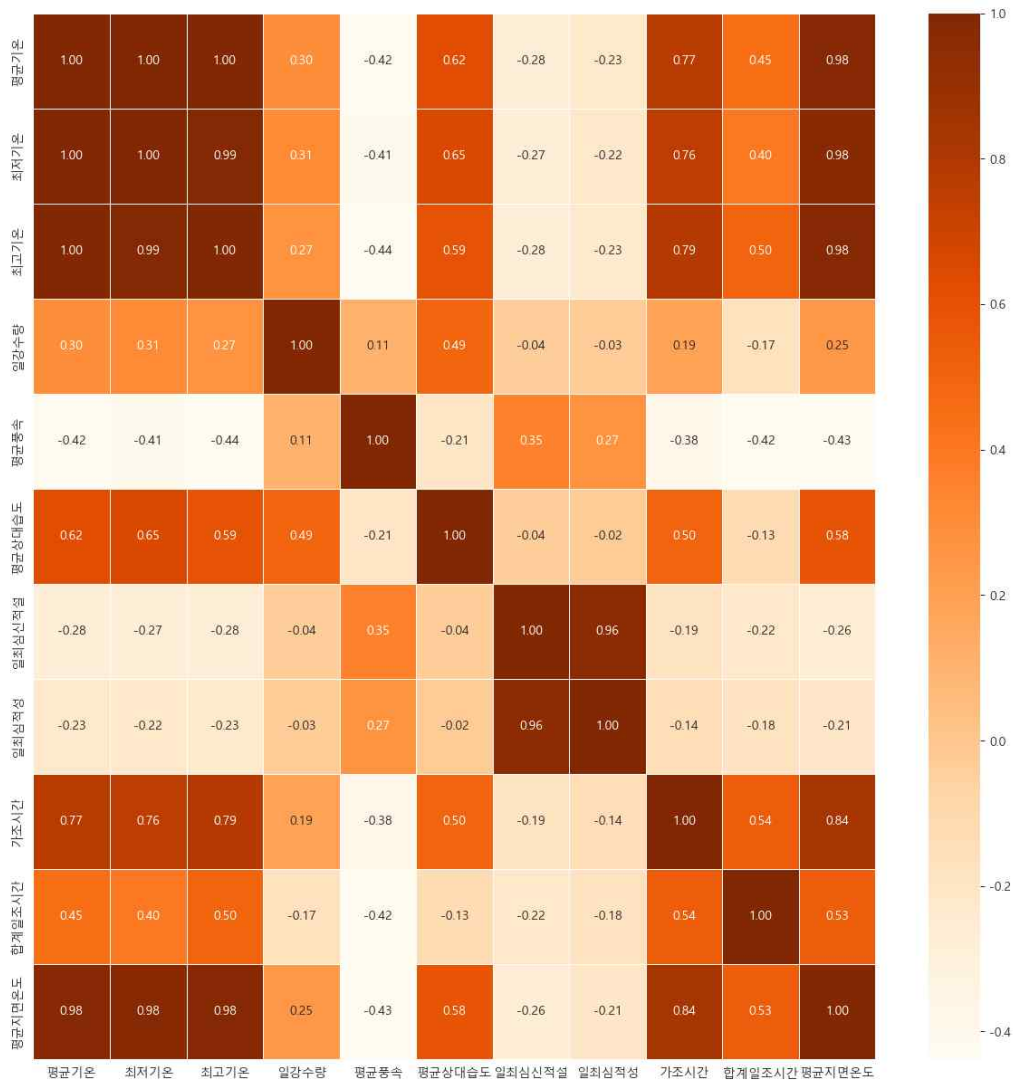


[그림 19] 지역별 평균지면온도(°C) 추이 시각화 결과

위 EDA 결과를 보면 제주시와 서귀포시에 따라 토지 및 기후 특징이 다르므로 기상 측정값에 차이를 보인다. 기후 정보는 감굴의 출하량 예측 시 가장 중요한 변수로 모델링 시 지역 특성을 반영하기 위해 출하량 비중에 따라 기상 측정값을 제주시와 서귀포시로 분류하여 가중평균을 적용해 모델에 반영하였다.

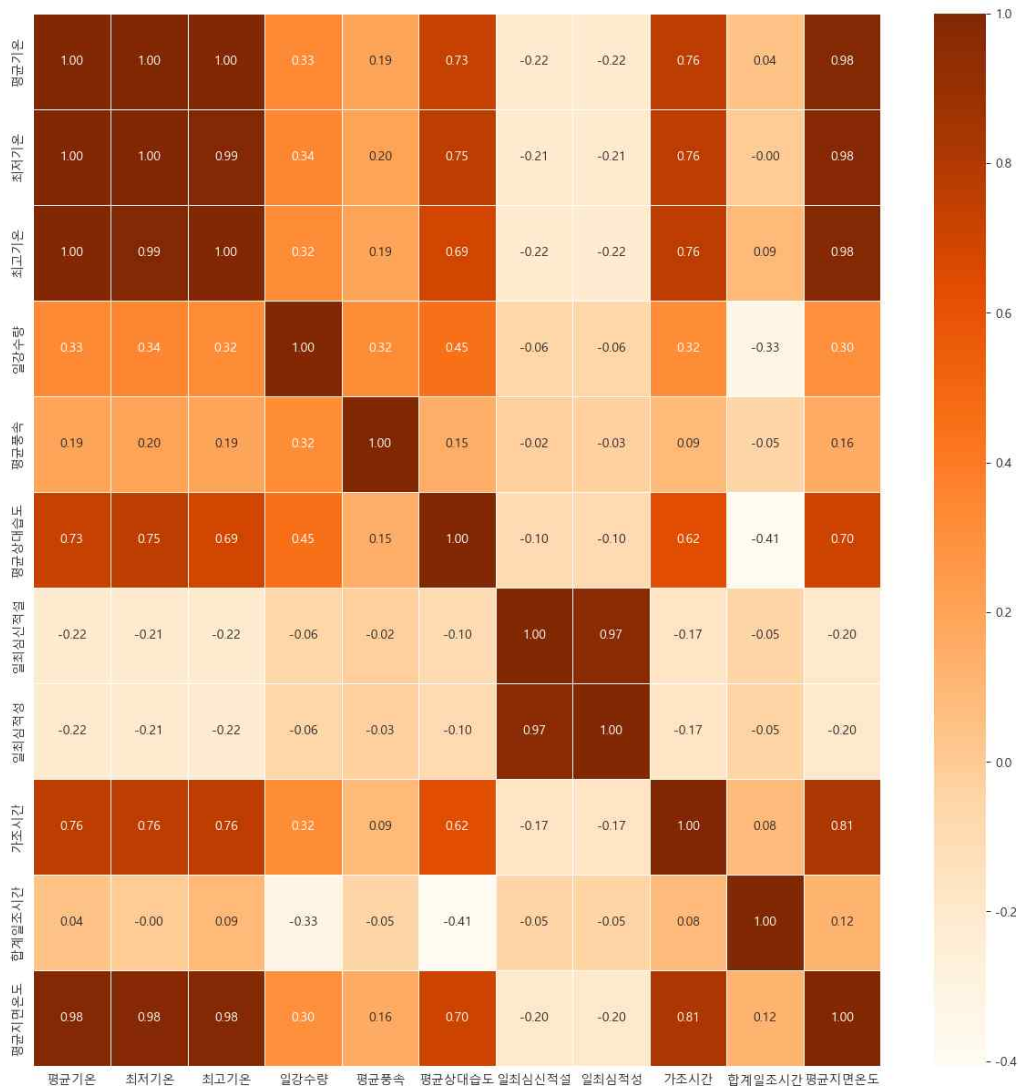
다음으로 진행한 EDA는 제주시와 서귀포시를 분류하여 기상 요인 간 상관관계를 파악하였다.

제주시 기상 요인 간 관계를 살펴보면 [그림20]과 같이 평균기온, 최저기온, 최고기온, 평균지면온도 사이에는 강한 양의 상관관계가 있으며, 평균풍속과 기온 사이에는 음의 상관관계, 일강수량과 기온 사이에는 약한 양의 상관관계를 보인다. 또한 일최심신적설과 일최심적설은 강한 양의 상관관계를 가지고 있으며, 평균풍속과는 약한 양의 상관관계를 보이는 것을 확인할 수 있었다.



[그림 20] 제주시 기상 요인 간 상관관계

서귀포시 기상 요인 간 관계를 살펴보면 [그림21]과 같이 평균기온, 최저기온, 최고기온 평균지면온도 사이에는 제주시와 같이 강한 양의 상관관계가 있으며, 평균풍속과 기온 사이에는 약한 양의 상관관계, 일강수량과 기온 사이에는 약한 양의 상관관계를 보인다. 또한 평균상대습도와 기온, 평균지면온도 사이에는 제주시보다 강한 양의 상관관계를 보였다.



[그림 21] 서귀포시 기상 요인 간 상관관계

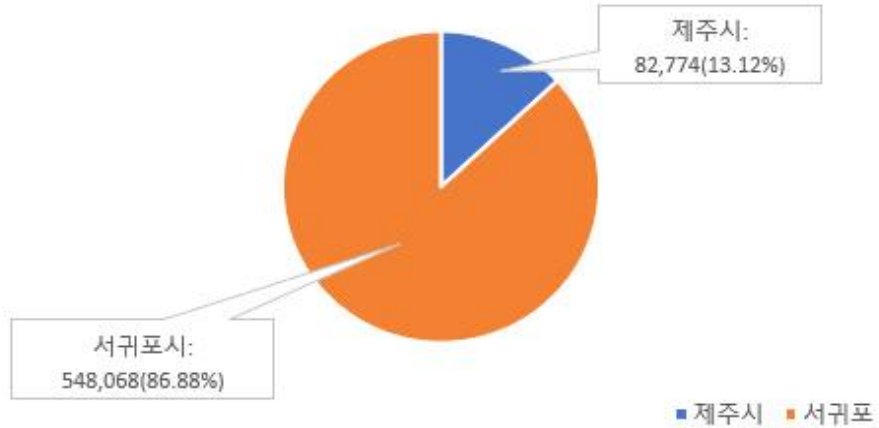
기상 요인 간에는 서로 상관성이 높으므로 상관관계가 높은 요인은 제거하여 비슷한 항목은 평균 요인 하나만을 모델링에 사용하였다.

4.1.2. 감귤 출하량 데이터 EDA

첫 번째로 진행한 EDA는 제주지역별로 제주시와 서귀포시의 감귤 1년간 출하량(2022년 기준)을 비교한 결과 [그림22]와 [표20]과 같이 출하 비율이 서귀포가 548,068톤(86.88%), 제주가 82,774톤(13.12%)으로 약 6.7배가량 차이가 나는 것으로 나타났다. 이는 대부분의 감귤이 서귀포시에서 출하되어 전국으로 판매되고

있는 것을 의미한다.

제주 지역별 감귤 출하량 분포



[그림 22] 제주 지역별 감귤 출하량 분포

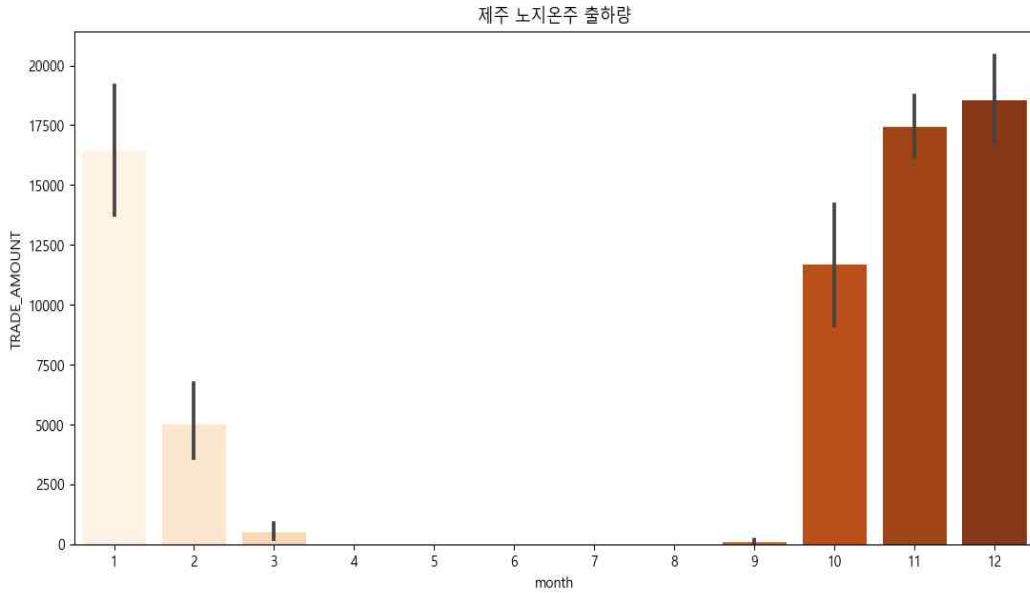
[표 20] 제주 지역별 감귤 출하량 분포

	제주시 출하량(톤)	서귀포시 출하량(톤)	제주시 비율(%)	서귀포시 비율(%)
하우스온주	18,071	220,725	7.57%	92.43%
노지온주	64,703	327,343	16.50%	83.50%
전체	82,774	548,068	13.12%	86.88%

이 결과는 출하량 예측 결과의 정확도를 높이기 위해 출하량 데이터와 날씨 데이터를 병합할 때 지역별 출하량 비중에 맞춰 기상 데이터의 측정값에 지역별 가중평균을 적용하는 데 활용된다.

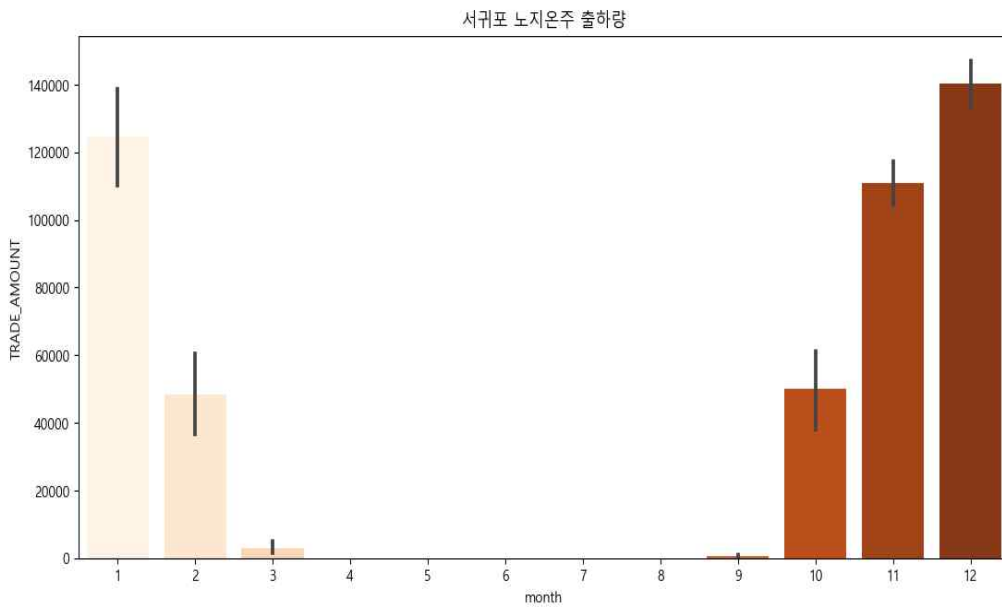
두 번째로 진행한 EDA는 감귤 품종별 출하량을 월별 집계하여 품종별 출하 시기를 파악하였다. 감귤 품종별로 출하 시기가 모두 다르므로 출하 시기를 모델에 반영함으로써 모델의 성능을 높이는 데 큰 요인으로 작용하였다. 제주시와 서귀포시를 구분하여 노지온주와 하우스온주의 출하 시기를 살펴보면 다음과 같다.

- 제주시 노지온주 출하량을 통해 알아본 출하시기: 10월~3월



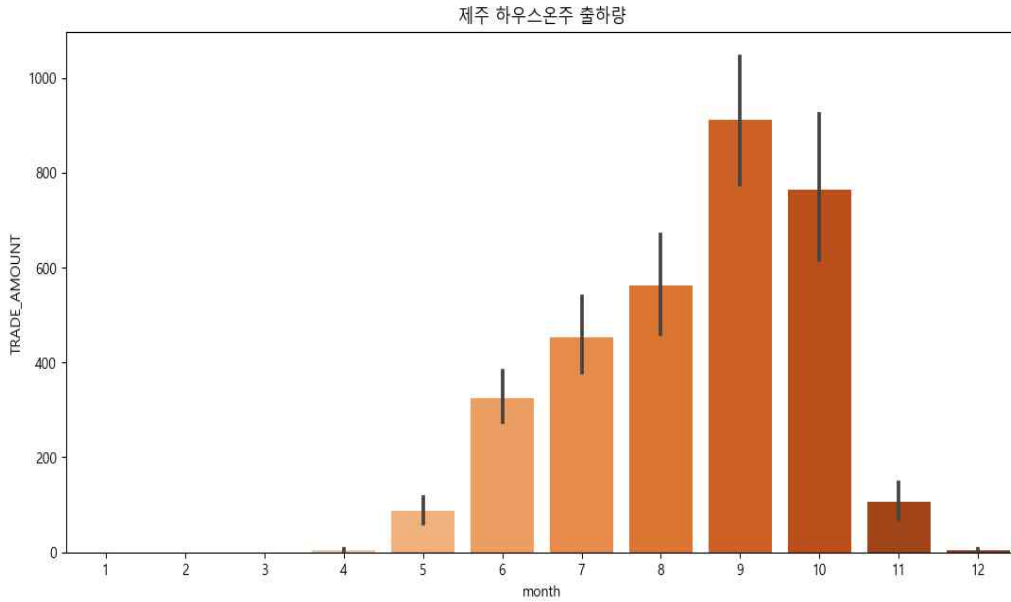
[그림 23] 월별 제주시 노지온주 출하량(톤)

- 월별 서귀포시 노지온주 출하량을 통해 알아본 출하시기: 10월~3월
- 서귀포시의 출하량이 제주시에 비해 절대적으로 높음. 거의 대부분의 노지온주는 서귀포시에서 출하한다는 것을 의미



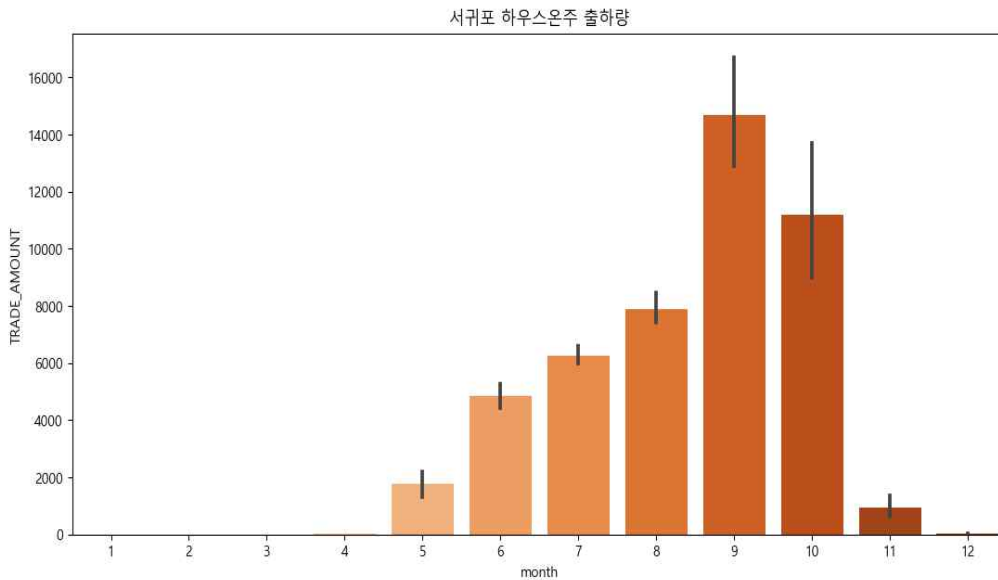
[그림 24] 월별 서귀포시 노지온주 출하량(톤)

- 월별 제주시 하우스온주 출하량을 통해 알아본 출하시기: 5월~11월



[그림 25] 월별 제주시 하우스온주 출하량(톤)

- 월별 서귀포시 하우스온주 출하량을 통해 알아본 출하시기: 5월~11월
- 서귀포시의 출하량이 제주시에 비해 절대적으로 높음. 거의 대부분의 하우스온주는 서귀포시에서 출하한다는 것을 의미



[그림 26] 월별 서귀포시 하우스온주 출하량(톤)

위 EDA 결과를 보면 노지온주의 경우 10월~3월, 하우스온주의 경우 5월~11월 사이에 대부분의 감귤이 출하된다는 것을 알 수 있다. 타과일 또한 품종별로 출

하 시기가 다른데 비슷한 시기에 출하되는 과일들끼리 도매시장 가격에 영향을 많이 주게 된다. 우리는 품종별 출하량 비중과 품종별 판매 거래량 비중을 함께 고려하여 감귤 품종별로 경쟁 과일 품목을 선정하였다.

4.1.3. 과일 도매가격 및 거래량 데이터 EDA

첫 번째로 진행한 EDA는 감귤 품종에 따른 지역별 거래량 비중을 알아보았다. 감귤 가격은 품종에 따라서도 차이가 크게 나지만 수도권 지역과 비수도권 지역의 도매시장에 따라서도 감귤 가격 차이가 크게 나는 것을 확인하였다.

그래서 이번 연구에서는 감귤 가격 및 거래량 데이터에서 도매시장명에 따라 지역별 9대 도매시장을 구분 후 수도권과 비수도권에 따라 나누는 라벨링 작업을 진행하였다. (제주감귤출하연합회, 2022)

수도권과 비수도권으로 분류된 라벨링 결과에 따라 감귤 품종별 거래량 비중을 살펴보면 [표21]과 같다.

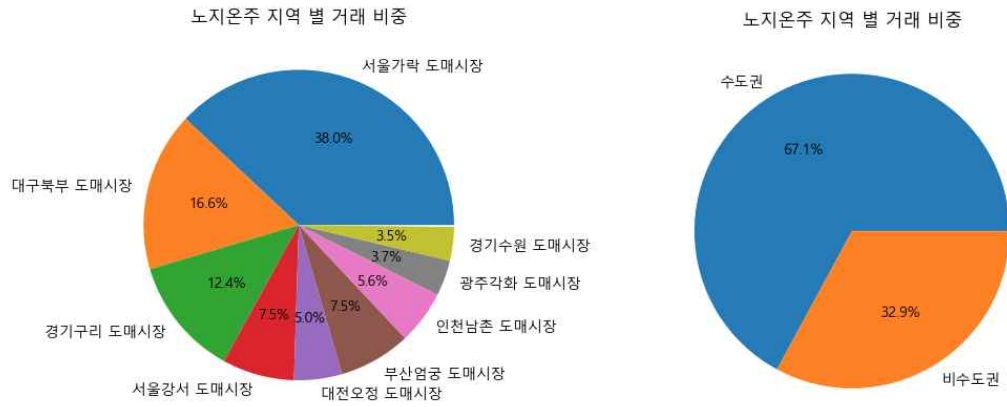
- 감귤 품종별 수도권 거래량 비중이 평균 71.83%정도 차지함

[표 21] 감귤 품종별 수도권/비수도권 거래량 비중

감귤 품종명	수도권 비중(%)	비수도권비중 (%, 1-수도권비중)
노지온주	0.671	0.329
하우스온주	0.668	0.332

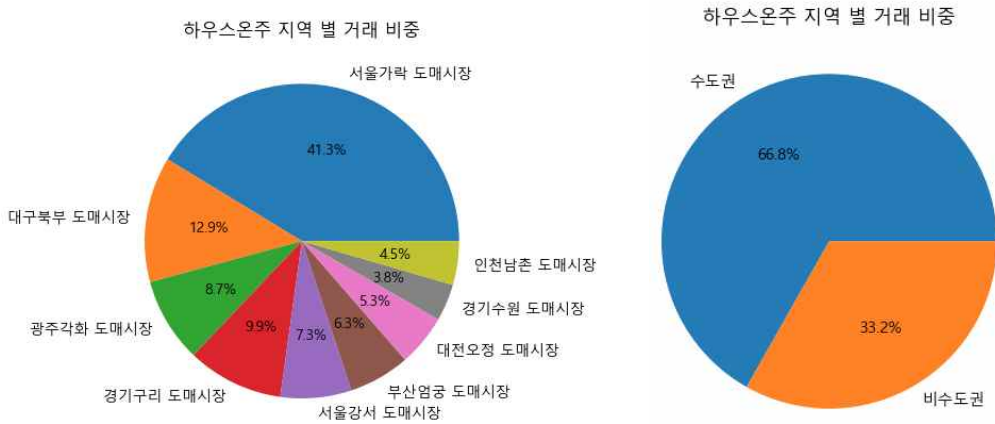
9대 도매시장별로 분류된 라벨링 결과에 따른 거래량 비중과 수도권과 비수도권으로 분류된 라벨링 결과에 따라 거래량 비중을 함께 시각화하여 감귤 품종별로 살펴보면 아래와 같다.

- 노지온주의 9대 도매시장별, 지역별 거래 비중



[그림 27] 노지온주의 9대 도매시장별, 지역별 거래 비중(%)

- 하우스온주 감귤의 수도권/비수도권 거래 비중



[그림 28] 하우스온주의 9대 도매시장별, 지역별 거래 비중(%)

위 결과를 보면 노지온주와 하우스온주 모두 서울가락 도매시장에서 약 40% 정도 거래되며 수도권 지역에서 거래되는 비중이 약 67%로 높은 것을 알 수 있다. 그 뒤로는 대구북부 도매시장, 경기구리 도매시장, 서울강서 도매시장, 광주각화 도매시장 순으로 거래가 일어나고 있는 것을 알 수 있다.

두 번째로 진행한 EDA는 수도권과 비수도권에 따른 과일 주거래 시기를 알아 보았다. 본 데이터에서 수집되는 과일 종류는 [표122]와 같이 16종이다.

[표 22] 과일별 가격 및 거래량 데이터에 수집된 과일 종류

번호	과일명	번호	과일명
1	노지온주	9	월동온주
2	단감	10	참외
3	딸기	11	천혜향
4	레드향	12	카라향
5	배	13	포도
6	사과	14	하우스온주
7	수박	15	한라봉
8	수입오렌지	16	황금향

위 과일에 대해 수도권과 비수도권을 구분하여 월별로 거래량을 비교하면 아래의 [표23], [표24]와 같이 나타났다.

[표 23] 수도권 지역의 과일별 거래량

(단위: KG)

월	노지온주	단감	딸기	레드향	배	사과	수박	수입오렌지
1	211758.0	82859	33914.20	3521.7	279787.5	253408.0	13189.6	41414.40
2	92171.0	54583	41587.20	2663.4	235597.5	176225.0	12805.6	73686.60
3	8820.5	26923	63249.32	236.7	72565.5	168581.0	18436.8	409167.00
4	3.0	4805	38875.80	19.2	66949.5	137340.0	54862.4	438608.16
5	0.0	550	14727.60	0.0	59782.5	109420.0	285048.0	308995.20
6	0.0	20	786.00	0.0	48015.0	88430.0	522104.0	78600.60
7	0.0	9	16.00	0.0	45060.0	71980.0	499224.0	29228.40
8	0.0	7	0.00	0.0	52723.5	117522.0	273866.4	35755.20
9	408.0	17600	2.20	0.0	386377.5	271661.0	50410.4	36023.40
10	108144.5	219859	134.80	0.0	121437.0	203723.2	25747.2	30560.40
11	206203.0	266795	5254.80	2.4	159607.5	197978.0	20960.0	24408.00
12	266599.0	115826	25744.60	1434.6	121390.5	179251.0	16240.8	27504.00
월	월동온주	참외	천혜향	카라향	포도	하우스온주	한라봉	황금향
1	92.7	901	4557.0	0.0	6610.00	0.0	3587.1	114.6
2	525.3	12055	5259.3	0.0	3954.00	0.0	3580.8	1.8
3	951.0	77728	3101.7	11.1	6144.00	0.0	1100.7	0.0
4	206.1	143143	1105.8	521.7	8572.00	14.7	389.4	0.0
5	0.0	334850	47.4	416.7	10984.00	1203.0	76.2	0.0
6	0.0	263000	0.0	3.0	9146.00	2695.5	0.0	0.0
7	0.0	240200	0.0	0.0	8690.00	2862.6	0.0	6.6
8	0.0	68943	0.0	0.0	31870.00	3420.6	0.0	23.1
9	0.0	33886	0.0	0.0	55620.60	7281.6	0.0	218.1
10	0.0	6144	0.0	0.0	37649.80	5118.9	3.0	260.7
11	0.0	466	99.0	0.0	13437.40	664.5	407.4	1927.5
12	69.9	630	1980.0	0.0	9483.02	56.1	2354.7	1710.0

[표 24] 비수도권 지역의 과일별 거래량

(단위: KG)

월	노지온주	단감	딸기	레드향	배	사과	수박	수입오렌지
1	197425.5	45508.0	18123.20	3690.3	220525.5	202723.4	5758.72	9309.60
2	88458.0	32599.8	22964.40	2369.1	179901.0	129180.0	7009.60	31867.20
3	7866.5	14558.0	33545.10	147.9	52222.5	114922.0	5500.00	267697.80
4	2.0	905.0	2948.80	14.4	5737.2	10316.0	541.60	253769.40
5	0.0	9.0	30.40	0.0	0.0	0.0	0.00	124194.60
6	0.0	190.0	0.00	0.0	0.0	0.0	0.00	20154.60
7	0.0	1.0	0.20	0.0	0.0	0.0	0.00	4456.80
8	0.0	0.0	0.00	0.0	0.0	0.0	12.80	5936.40
9	513.0	9494.0	1.00	0.0	87970.5	54428.0	6552.80	6336.36
10	112873.0	159528.0	16.48	0.0	104755.5	144492.4	9227.20	5063.40
11	195836.0	389281.0	2757.20	4.2	125535.0	162347.0	6927.20	4084.20
12	249485.5	65910.9	11533.80	1513.2	102252.0	133777.3	4694.40	3681.00
월	월동온주	참외	천혜향	카라향	포도	하우스온주	한라봉	황금향
1	61.2	398.0	5097.6	0.0	474.8	0.0	3765.6	45.0
2	325.5	6586.0	5717.7	0.0	207.4	0.0	3246.9	1.5
3	408.6	36551.4	2463.3	4.5	213.8	0.0	663.3	0.0
4	74.7	4602.0	842.4	501.0	87.2	4.5	83.7	0.0
5	0.0	0.0	24.6	288.6	0.0	818.4	18.3	0.0
6	0.0	0.0	0.0	2.7	0.0	2185.2	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	2650.2	0.0	1.5
8	0.0	0.0	0.0	0.0	0.0	3026.1	0.0	6.6
9	0.0	2404.0	0.0	0.0	10223.4	7967.7	0.0	232.8
10	0.0	2118.0	0.0	0.0	22357.8	3571.8	0.0	318.6
11	0.0	136.0	91.5	0.0	2629.0	279.6	452.1	1609.5
12	34.2	22.0	1998.3	0.0	1102.2	32.4	2431.5	1210.5

감귤별 거래량이 가장 많은 시기(4개월)와 타과일 거래량이 가장 많은 시기가 3개월 이상 겹치는 경우를 추출한 결과는 [표25]와 같다. 노지온주의 경우 수도권은 3종류의 과일이 경쟁 과일로 선정되었고, 비수도권은 5종류의 과일이 경쟁 과일로 선정되었다. 하우스온주의 경우 수도권 경쟁 과일로 1종류의 과일이 선정되었다.

[표 25] 감귤 품종별 지역별 경쟁 과일 목록

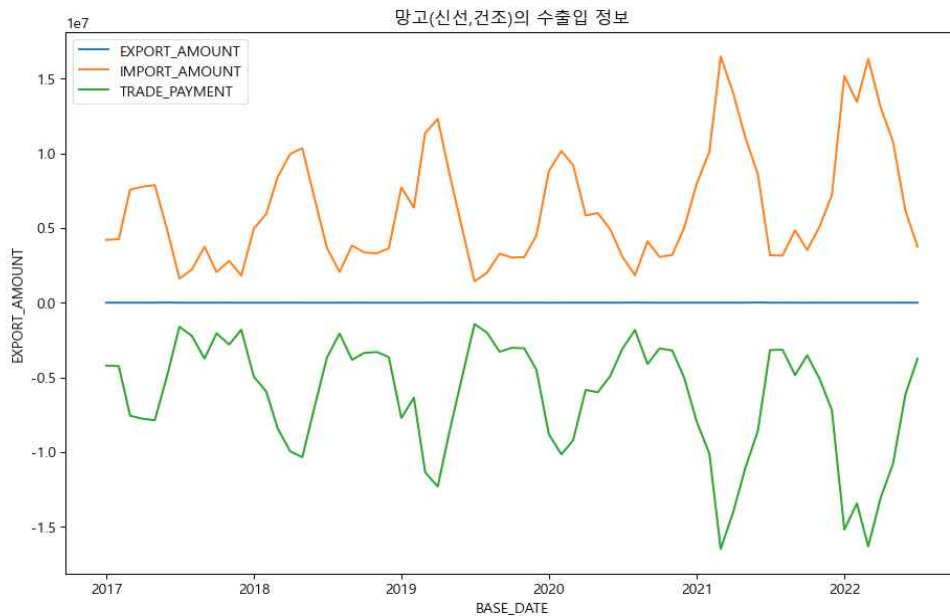
	수도권 경쟁 과일	비수도권 경쟁 과일
노지온주	단감, 사과, 황금향	단감, 배, 사과, 포도, 황금향
하우스온주	포도	-

감귤의 가격을 예측할 때 타과일/타품종의 모든 경쟁 과일을 고려한다면 데이터 차원이 증가할 뿐만 아니라 예측에도 도움이 되지 않음을 확인하여 각 품종의 가격을 예측하는 데 영향을 많이 주는 경쟁 과일을 위와 같은 기준으로 정의하여 가격 예측 시에는 해당 경쟁 과일의 데이터(최소가/평균가/최고가)만 사용하였다. 따라서 가격 예측 모델의 입력차원은 수도권/비수도권, 노지온주/하우스 온주에 따라 경쟁 과일 수가 달라지므로 입력 차원 수의 차이가 발생한다.

4.1.4. 과일 수입/수출 데이터 EDA

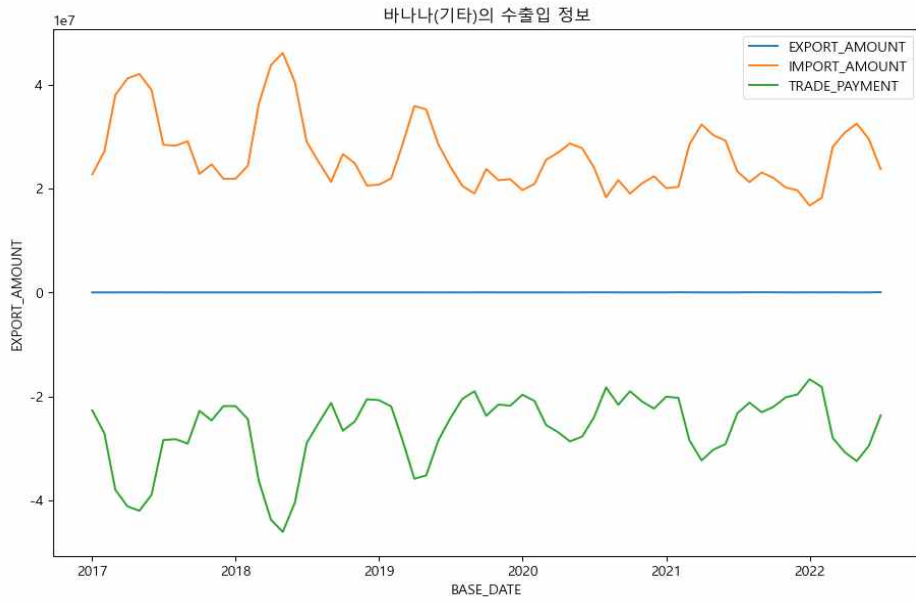
첫 번째로 진행한 EDA는 과일별로 기간에 따른 수입량, 수출량, 무역수지 변화 추이로 결과는 아래와 같다. 이때 TRADE_AMOUNT(무역수지)는 상품의 수출입 거래에서 생기는 국제수지로서 상품의 수출액(EXPORT_AMOUNT)과 수입액(IMPORT_AMOUNT)의 차액을 의미한다.

- 망고(신선,건조): 수입금액 > 수출금액(≈ 0), 최근 수입과일 수요량 증가 추세



[그림 29] 망고(신선, 건조)의 수출입 정보(달러)

- 바나나(기타): 수입금액 > 수출금액(≈ 0), 수입과일



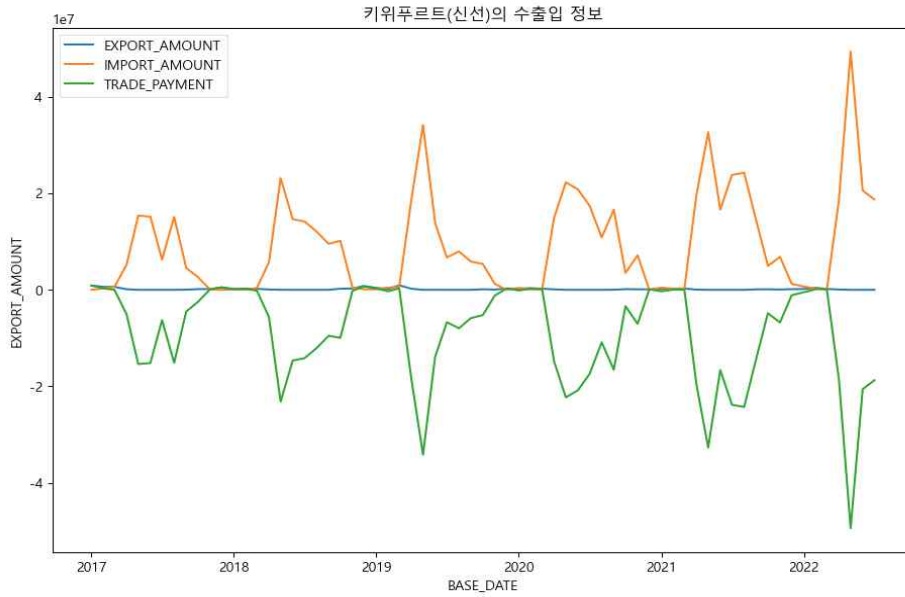
[그림 30] 바나나(기타)의 수출입 정보(달러)

- 오렌지(신선,건조): 수입금액 > 수출금액(≈ 0), 수입과일



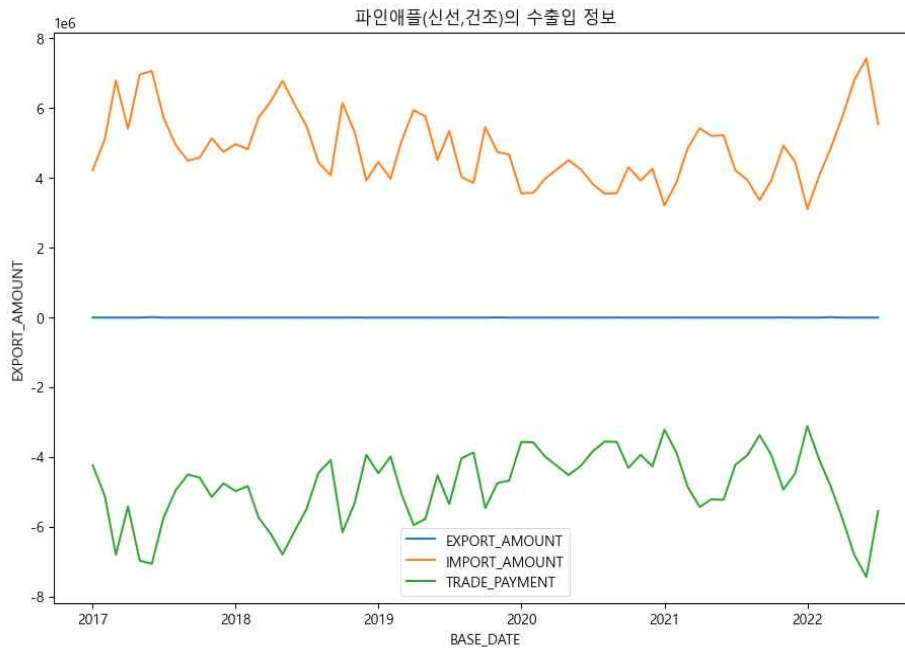
[그림 31] 오렌지(신선,건조)의 수출입 정보(달러)

- 키위푸르트(신선): 수입금액 > 수출금액(≈ 0), 최근 수입과일 수요량 증가 추세



[그림 32] 키위푸르트(신선)의 수출입 정보(달러)

- 파인애플(신선,건조): 수입금액 > 수출금액(≈ 0), 수입과일



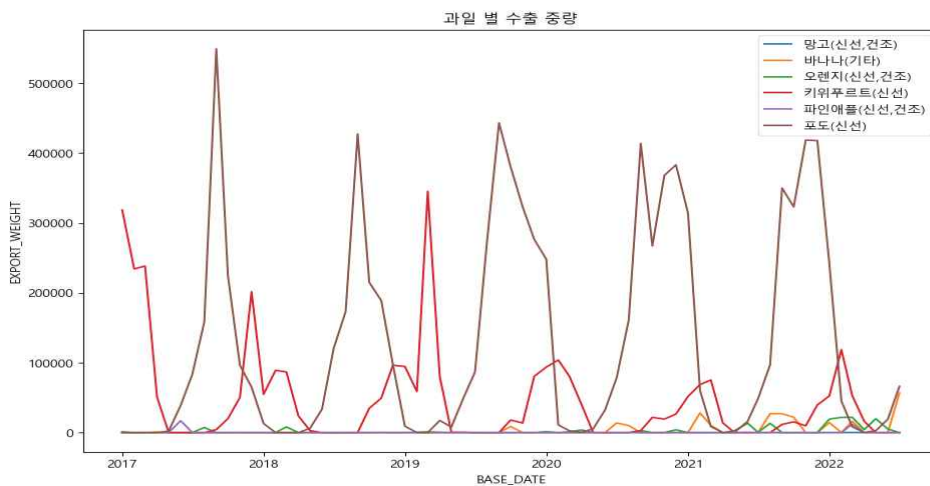
[그림 33] 파인애플(신선,건조)의 수출입 정보(달러)

- 포도(신선): 수입금액 > 수출금액, 수출금액이 0에 가까운 값이 아닌 최근 수출금액이 점차 증가하는 추세를 보이며, 자국 생산이 어느 정도 가능한 것으로 보임



[그림 34] 포도(신선)의 수출입 정보(달러)

두 번째로 진행한 EDA는 과일별 수출 중량을 비교하였다. [그림35]를 보면 과일별 수출 중량은 포도(신선)가 가장 높았으며 그다음으로 키위푸르트(신선)가 높은 비중을 차지하고 있었다. 그리고 최근 들어 수입 과일의 수출 중량이 전체적으로 증가함을 알 수 있으며, 코로나19 발생 이후 키위푸르트(신선)의 수출 중량이 급격히 낮아진 것을 확인할 수 있었다.



[그림 35] 과일별 수출 중량(KG)

4.2. 실험 모형 및 예측 결과

4.2.1. 감귤 출하량 예측 모형

본 연구에서 개발한 출하량 예측 모형은 기상 데이터를 입력으로 받아 노지온주와 하우스온주의 출하량을 예측하는 모형이다. 기상 데이터의 경우 제주시와 서귀포로 지역을 구분하고 지역별 출하량 비중에 맞춰 날씨 정보를 가중 평균하여 사용하였다. 최종적으로 기상 데이터와 출하량 데이터를 병합하여 만들어진 데이터 마트에 대한 요약은 [표26], [표27]과 같다.

- 입력 데이터: [날씨정보]평균기온(°C), 일강수량(mm), 평균풍속(m/s), 합계일조시간(hr), 일최심신적설(cm), [출하시기]월
- 출력 데이터: [출하량정보]노지온주 출하량(톤), 하우스온주 출하량(톤)

[표 26] 노지온주 출하량 예측 데이터 요약 정보

노지온주						
	TRADE AMOUNT	평균기온	일강수량	평균풍속	합계일조시간	일최심신적설
count	314	314	314	314	314	314
mean	46166.1	17.05464	5.085688	1.93604	6.010663	0.014867
std	64422.91	7.034975	8.358876	0.414506	1.961282	0.094648
min	0	2.336408	0	1.127427	1.073335	0
25%	0	11.00909	0.231233	1.646493	4.537033	0
50%	44.6225	17.28312	1.926968	1.904889	5.902072	0
75%	103763	23.04762	5.462699	2.124926	7.29528	0
max	206764.3	29.72814	70.04342	3.73372	11.24021	1.026353

[표 27] 하우스온주 출하량 예측 데이터 요약 정보

하우스온주						
	TRADE AMOUNT	평균기온	일강수량	평균풍속	합계일조시간	일최심신적설
count	285	285	285	285	285	285
mean	4659.011	17.61028	5.441228	1.85186	6.036419	0.008481
std	5961.44	6.940725	8.841607	0.425694	1.978976	0.061791
min	0	2.317899	0	1.217105	0.927117	0
25%	0	11.76516	0.23487	1.559996	4.662256	0
50%	2226.942	17.83656	2.043886	1.788836	5.931748	0
75%	7355.826	23.45628	5.98521	2.035832	7.314845	0
max	35133.27	29.69205	71.55372	3.671496	11.26334	0.767983

출하량 예측을 위한 학습 데이터는 학습(Train):평가(Test)=70:30 비율로 구성하였으며, 데이터 정규화 작업으로는 MinMaxScaling(정규화) 작업을 진행했다. 먼저 학습 데이터에 스케일러(Scaler)를 피팅(fitting)한 후 해당 스케일러를 사용하여 학습 데이터와 평가 데이터를 변환하여 [그림36]과 같이 정규화된 데이터를 추출하였다.

평균기간	일강수량	평균풍속	합계일조시간	일최심신적심	month_1	month_2	month_3	month_4	month_5	month_6	month_7	month_8	month_9	month_10	month_11	month_12	TRADE_AMOUNT
2017-01-01	0.268001	0.000000	0.000000	0.513529	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.149015
2017-01-08	0.322779	0.097730	0.380388	0.253513	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.979353
2017-01-15	0.116094	0.000000	0.390908	0.505445	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.827110
2017-01-22	0.099919	0.003858	0.426285	0.383493	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.764459
2017-01-29	0.162080	0.047584	0.307757	0.545863	0.000000	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.363336
2017-02-05	0.191812	0.048834	0.259402	0.468555	0.000000	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.254539
2017-02-12	0.067489	0.006990	0.532487	0.476413	0.252941	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.134873
2017-02-19	0.257011	0.009807	0.395244	0.725333	0.000000	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.065235
2017-02-26	0.209523	0.170392	0.451121	0.475029	0.000000	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.018234
2017-03-05	0.234145	0.000427	0.338569	0.698105	0.000000	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.005642

[그림 36] 노지온주 출하량 데이터 입/출력 변수 정규화

예측 모형은 '2.2. 연구방법'에서 언급했던 심층신경망 기반의 시계열 예측 모델로 LSTM, Seq2Seq, Attention 모델을 선정하여 예측력에 초점을 두어 실험을 진행했으며, 설정한 주요 모수는 [그림3]과 같다. 주요 모수 이외에도 모델 최적화를 위해 Model Capacity를 조정하는 등의 작업을 진행하였다. 비교 모델은 설명력에 초점을 둔 머신러닝 모델로 XGBoost Regressor와 Random Forest Regressor를 사용하였다.

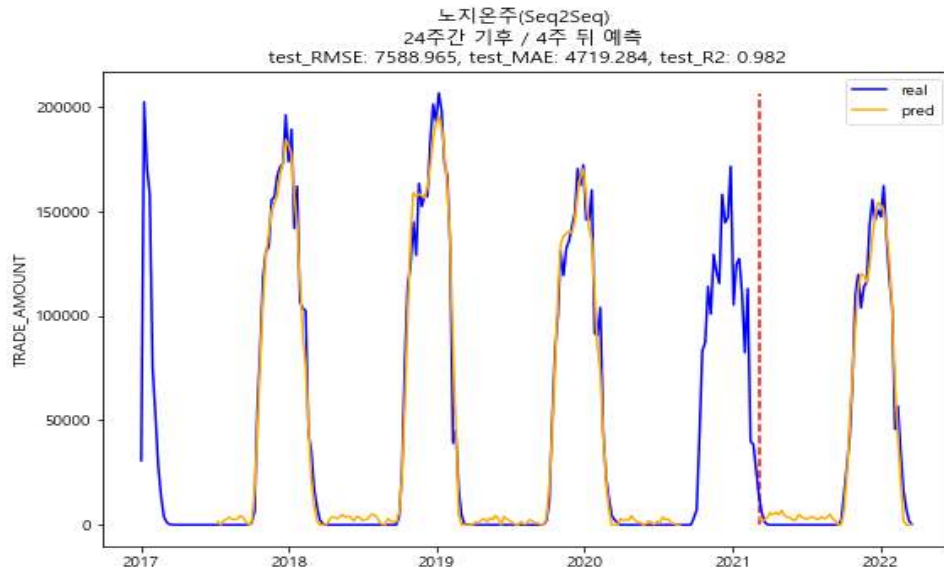
위에서 설정된 모델을 통해 노지온주에 대해 윈도우 사이즈별로 출하량을 예측하는 실험을 진행한 결과는 아래와 같다. 출하량 예측 모형을 구성할 때 시계열 모델 3개, 윈도우 사이즈는 입력 기간 20주(5개월), 24주(6개월), 출력 기간 4주(1개월), 8주(2개월), 12주(3개월)로 설정하여 총 18가지 경우에 대해 비교하는 실험을 진행하였다.

첫 번째 실험 결과는 [표28], [그림37]과 같이 입력을 20주(5개월)와 24주(6개월)로 하고 출력을 4주(1개월)로 진행한 경우로 Seq2Seq 모델이 가장 예측 성능

이 우수하였으며, LSTM과 Attention 모델은 비슷한 결과를 보였다. 입력 기간은 5개월인 경우보다 6개월로 설정했을 때 결과가 더 좋게 나타났다.

[표 28] 4주(1개월) 출하량 예측 결과: 입력-20주(5개월) / 24주(6개월)

노지온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	4	XGBoost	21674.91	15412.53	0.861
		Random Forest	21633.36	15772.77	0.842
		LSTM	16883.22	9282.275	0.908
		Seq2Seq	9829.492	7086.155	0.969
		Attention	16690.52	9284.741	0.91
24	4	XGBoost	23737.49	19660.89	0.836
		Random Forest	21709.6	16726.56	0.833
		LSTM	12018.15	7194.405	0.954
		Seq2Seq	7588.965	4719.284	0.982
		Attention	12046.73	6698.222	0.953



[그림 37] 노지온주 4주(1개월) 뒤 출하량 예측 Best result

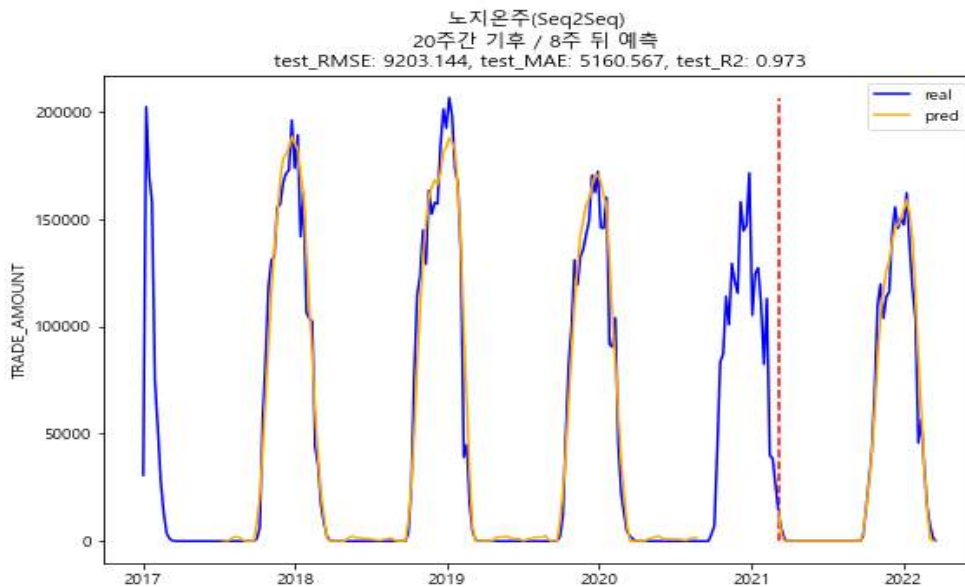
두 번째 실험 결과는 [표29], [그림38]과 같이 입력을 20주(5개월)과 24주(6개월)로 하고 출력을 8주(2개월)로 진행한 경우로 Seq2Seq 모델이 가장 예측 성능

이 우수하였으며 LSTM과 Attention 모델은 비슷한 결과를 보였다. 입력 기간은 4주 예측과 달리 5개월을 사용하였을 때 결과가 더 좋게 나타났다.

- 8주(2달) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

[표 29] 8주(2개월) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

노지온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	8	XGBoost	21466.91	15110.56	0.863
		Random Forest	21463.79	17877.64	0.843
		LSTM	16238.99	10276.05	0.917
		Seq2Seq	9203.144	5160.567	0.973
		Attention	16100.61	10002.61	0.919
24	8	XGBoost	20690.58	15362.93	0.897
		Random Forest	25274.75	21026.97	0.796
		LSTM	14729.08	9284.098	0.934
		Seq2Seq	9995.691	5301.743	0.97
		Attention	16248.78	10651.18	0.92

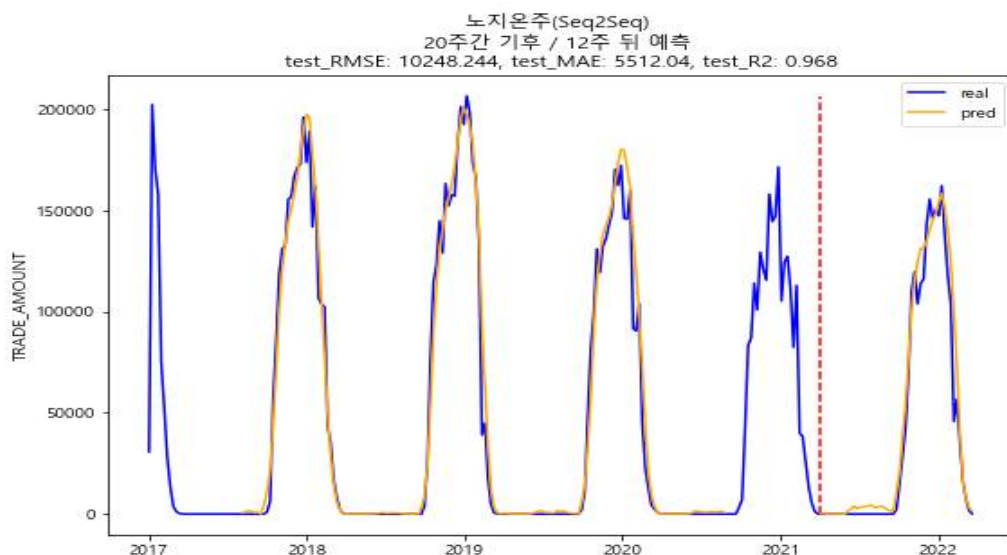


[그림 38] 노지온주 8주(2개월) 뒤 출하량 예측 Best result

세 번째 실험 결과는 [표30], [그림39]와 같이 입력을 20주(5개월)과 24주(6개월)로 하고 출력을 12주(3개월)로 진행한 경우로 Seq2Seq 모델이 가장 예측 성능이 우수하였으며 8주 예측 결과와 비교 시 크게 성능이 저하되지 않은 것을 확인할 수 있었다. 입력 기간은 5개월을 사용하였을 때 결과가 더 좋게 나타났다.

[표 30] 12주(3달) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

노지온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	12	XGBoost	22020.42	17601.8	0.858
		Random Forest	23216.11	18192.94	0.848
		LSTM	15928.65	10550.82	0.922
		Seq2Seq	10248.24	5512.04	0.968
		Attention	15689.83	10062.44	0.925
24	12	XGBoost	23192.73	19576.06	0.846
		Random Forest	20785.31	16499.71	0.856
		LSTM	13256.78	8753.946	0.948
		Seq2Seq	9705.564	5637.771	0.952
		Attention	17660.77	11183.22	0.908



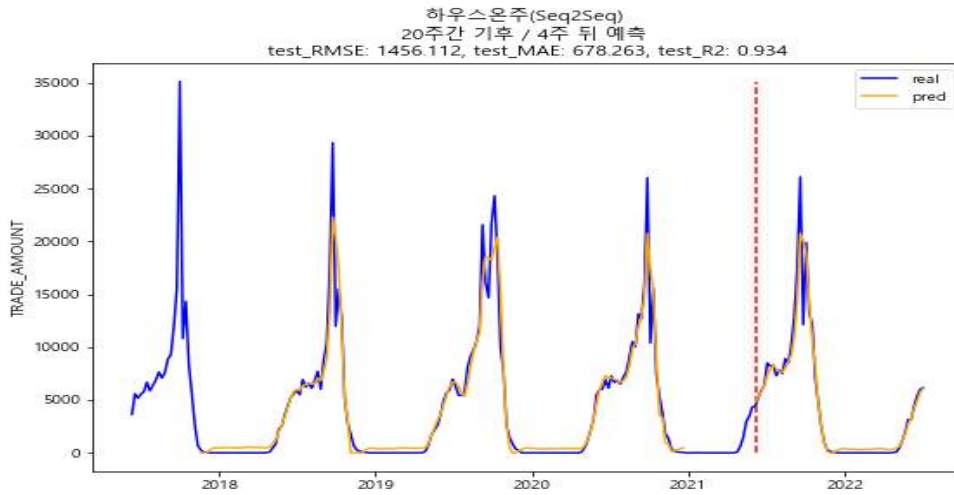
[그림 39] 노지온주 12주(3달) 뒤 출하량 예측 Best result

위에서 설정된 모델을 통해 하우스온주에 대해 윈도우 사이즈별로 출하량을 예측하는 실험을 진행한 결과는 아래와 같다. 출하량 예측 모델을 구성할 때 시계열 모델 3개, 윈도우 사이즈는 노지온주와 동일하게 입력 기간은 20주(5개월), 24주(6개월), 출력 기간은 4주(1개월), 8주(2개월), 12주(3개월)로 설정하여 총 18가지 경우에 대해 비교하는 실험을 진행하였다.

첫 번째 실험 결과는 [표31], [그림40]과 같이 입력을 20주(5개월)과 24주(6개월)로 하고 출력을 4주(1개월)로 진행한 경우로 Seq2Seq 모델이 가장 예측 성능이 우수하였으며, 8주와 12주를 예측할 때보다 4주를 예측했을 때 성능이 가장 우수하게 나타났다. 입력 기간으로 5개월과 6개월을 사용하였을 때 적합도 관점에서는 비슷하였으나, 예측력 관점에서 5개월을 입력으로 사용하였을 때 결과가 더 좋게 나타났다.

[표 31] 4주(1달) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

하우스온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	4	XGBoost	1446.358	834.8222	0.901
		Random Forest	1961.704	1377.266	0.854
		LSTM	1705.886	836.651	0.909
		Seq2Seq	1456.112	678.263	0.934
		Attention	1823.807	1207.776	0.896
24	4	XGBoost	1498.594	1028.109	0.915
		Random Forest	1297.032	798.4299	0.916
		LSTM	1931.802	1063.717	0.892
		Seq2Seq	1504.24	694.268	0.934
		Attention	2042.362	1304.077	0.879

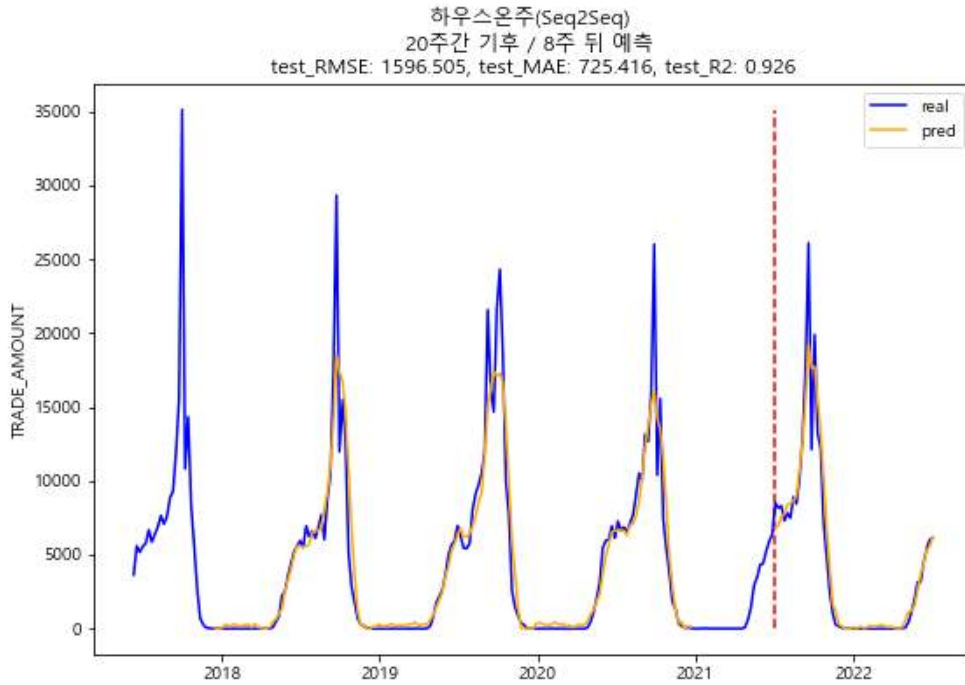


[그림 40] 하우스온주 4주(1개월) 뒤 출하량 예측 Best result

두 번째 실험 결과는 입력을 [표32], [그림41]과 같이 20주(5개월)과 24주(6개월)로 하고 출력을 8주(2개월)로 진행한 경우로 Seq2Seq 모델이 가장 예측 성능이 우수하였으며 LSTM과 Attention 모델은 비슷한 결과를 보였다. 입력 기간으로 5개월과 6개월을 사용하였을 때 적합도 관점에서는 비슷하였으나, 예측력 관점에서 6개월을 입력으로 사용하였을 때 결과가 더 좋게 나타났음

[표 32] 8주(2달) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

하우스온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	8	XGBoost	1670.411	914.2433	0.905
		Random Forest	1855.641	990.2088	0.897
		LSTM	1804.308	1027.484	0.905
		Seq2Seq	1596.505	725.416	0.926
		Attention	1867.702	1230.289	0.899
24	8	XGBoost	1245.592	744.9801	0.919
		Random Forest	1962.501	916.3924	0.908
		LSTM	1879.916	1050.049	0.904
		Seq2Seq	1628.131	750.461	0.928
		Attention	1904.82	1104.848	0.901

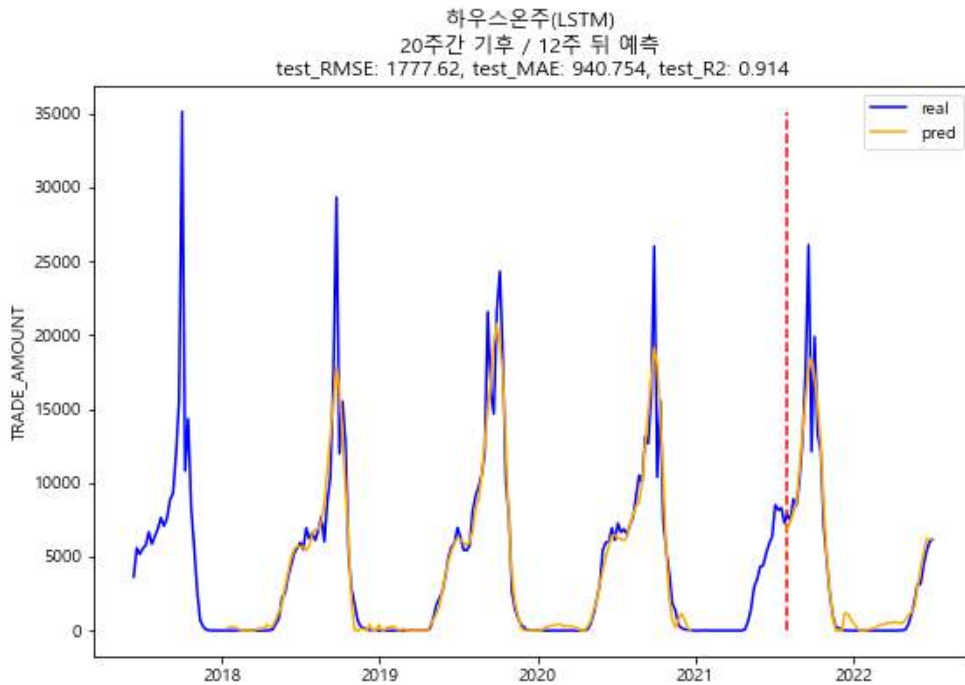


[그림 41] 하우스온주 8주(2개월) 뒤 출하량 예측 Best result

세 번째 실험 결과는 [표33], [그림42]와 같이 입력을 20주(5개월)과 24주(6개월)로 하고 출력을 12주(3개월)로 진행한 경우로 LSTM 모델이 가장 예측 성능 및 적합도 면에서 우수하였으며, 그다음으로 Seq2Seq과 Attention 모델이 적합하다고 나타났다. 모델에 따라 입력 기간이 5개월과 6개월을 사용할 때의 성능 차가 달랐으며, 가장 성능이 좋은 LSTM 모델에 대해서는 6개월을 입력으로 사용할 때 결과가 좋게 나타났다.

[표 33] 12주(3달) 출하량 예측 결과: 입력-20주(5개월)/24주(6개월)

하우스온주					
Input term	Output term	Model	RMSE	MAE	R2-Score
20	12	XGBoost	1648.055	1137.544	0.896
		Random Forest	1946.653	800.9064	0.881
		LSTM	1777.62	940.754	0.914
		Seq2Seq	2088.473	995.485	0.881
		Attention	1914.968	1143.975	0.9
24	12	XGBoost	1215.99	788.8854	0.887
		Random Forest	1936.218	1063.0851	0.865
		LSTM	1811.865	897.501	0.916
		Seq2Seq	2024.146	851.005	0.895
		Attention	2038.517	1310.087	0.894



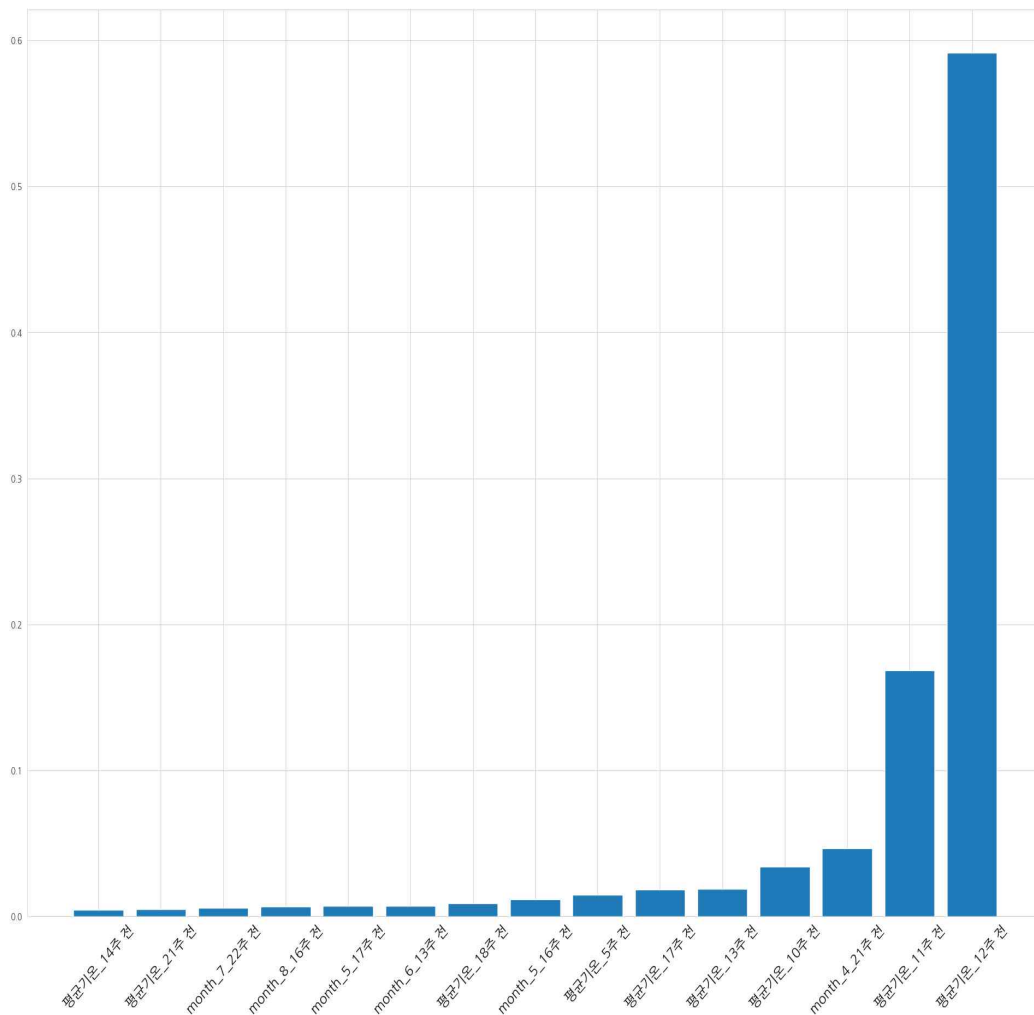
[그림 42] 하우스온주 12주(3개월) 뒤 출하량 예측 Best result

이를 통해 본 연구에서는 감귤의 출하량 예측 시 설명력에 초점을 둔 머신러닝 모형보다 예측력 기반의 심층신경망 모형의 성능이 좋게 나타나는 것을 확인

할 수 있었다.

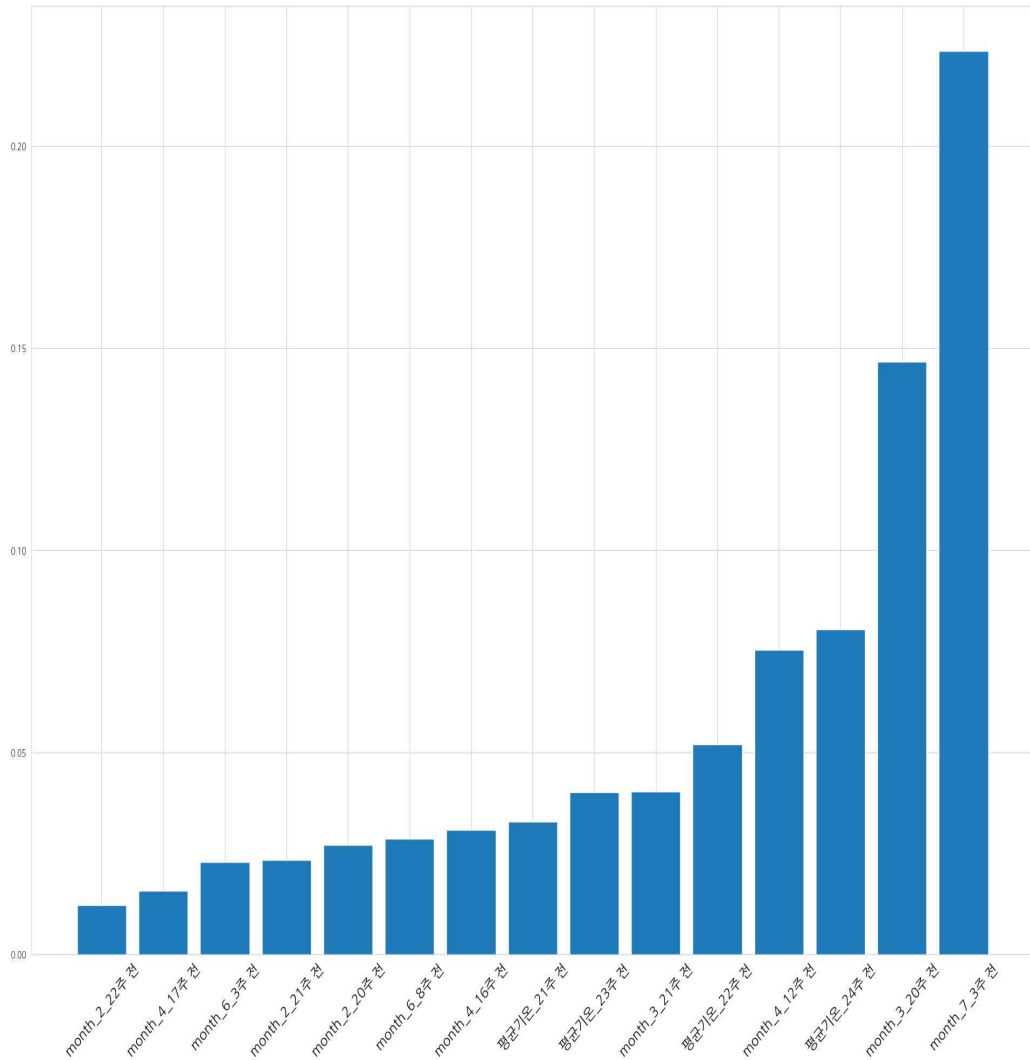
다음으로 설명력 기반의 모형을 통해 변수 중요도(Feature Importance)를 알아보았다. 감귤 출하량 예측 시 설명력 기반의 모형은 Random Forest보다 XGBoost가 성능이 우수하게 나타났으며, 노지온주와 하우스온주 모두 입력 기간을 24주(6개월), 출력 기간을 8주(2개월)로 설정하였을 때 R2-Score가 가장 높게 나타났다.

노지온주 출하량 예측 시 가장 우수한 성능을 보인 모형의 변수 중요도를 15위까지 출력해 보면 [그림43]과 같이 평균기온이 가장 큰 영향력을 보였으며 그 뒤로는 출하 시기 월 정보가 영향력 있는 요인으로 나타났다.



[그림 43] 머신러닝 기반 노지온주 출하량 예측 Best Model의 변수 중요도

하우스온주 출하량 예측 시 가장 우수한 성능을 보인 모형의 변수 중요도를 15위까지 출력해 보면 [그림44]와 같이 출하 시기 월 정보가 가장 큰 영향력을 보였으며 그 뒤로는 평균기온이 영향력 있는 요인으로 나타났다.



[그림 44] 머신러닝 기반 하우스온주 출하량 예측 Best Model의 변수 중요도

4.2.2. 감귤 가격 예측 모형

본 연구에서 개발한 가격 예측 모형은 감귤 출하량 데이터, 공휴일 데이터, 경쟁 과일의 가격 데이터, 감귤 가격 및 거래량 데이터를 입력으로 받아 노지온주와 하우스온주의 제주지역 최저가, 평균가, 최고가를 예측하는 모형이다.

이 실험을 위해 사용된 학습 데이터는 감귤 거래 특성을 반영하여 기존 수집하고 가공한 데이터에서 아래와 같은 처리를 추가하였다.

- 농가에서는 감귤이 출하되면 바로 판매하지 않고 창고에 보관하다가 판매하는 경우가 많다. 그래서 출하량 데이터에서 이전 4주간의 출하량의 합계를 집계하여 PAST_SHIPMENT(과거누적출하량) 변수를 파생 변수로 추가 생성
- 공휴일로 인해 소비자가 감귤을 구매할 때도 공휴일 당일이 아닌 미리 사전에 과일을 미리 구매하는 경우가 많으므로 공휴일 데이터에 공휴일 기준일 이전 일주일에 대해 모두 1로 라벨링을 진행하여 모델 학습 시 반영
- 경쟁 과일의 가격 데이터는 [표25]와 같이 감귤 품종별(노지온주, 하우스온주), 지역별(수도권, 비수도권)에 따라 선정된 경쟁 과일에 대해 최소가, 평균가, 최대가를 집계하여 추출

최종적으로 위 데이터들을 병합하여 만들어진 데이터 마트에 대한 요약은 [표 34]~[표45]와 같다.

- 입력 데이터: 감귤 가격(최저가, 평균가, 최고가)(원), 감귤 총 거래량(KG), 출하량(톤), 과거 누적 출하량(톤), 경쟁 과일 가격(최저가, 평균가, 최고가)(원)
 - 출력 데이터: 감귤(노지온주, 하우스온주)의 최저가(원), 평균가(원), 최고가(원)
- [표 34] 수도권 노지온주 최저가 예측 데이터

수도권_노지온주_PRICE_MIN							
	PRICE_MIN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MIN	사과_PRICE_MIN	황금향_PRICE_MIN
count	313	313	313	313	313	313	313
mean	1337.366	4131.313	46215.15	178726.3	2985.324	3899.892	4628.295
std	1439.302	5700.7	64520.19	247977.2	2580.264	1611.219	5453.821
min	0	0	0	0	0	0	0
25%	0	0	0	0	0	3000	0
50%	0	0	24.845	2069.9	3125	3875	0
75%	2625	10288	103767.5	401449.6	4857.143	4975	8400
max	4500	17967.5	206764.3	799071.8	9357.143	8600	30000

[표 35] 수도권 노지온주 평균가 예측 데이터

수도권_노지온주_PRICE_MEAN							
	PRICE_MEAN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MEAN	사과_PRICE_MEAN	황금향_PRICE_MEAN
count	313	313	313	313	313	313	313
mean	3777.327	4131.313	46215.15	178726.3	13106.42	24264.46	6839.874
std	4032.567	5700.7	64520.19	247977.2	10334.9	6027.534	7694.534
min	0	0	0	0	0	12101.25	0
25%	0	0	0	0	0	19866.67	0
50%	0	0	24.845	2069.9	15484.38	23500	0
75%	7537.5	10288	103767.5	401449.6	20938.54	28252.08	12379.17
max	11486.34	17967.5	206764.3	799071.8	38020	39683.33	32000

[표 36] 수도권 노지온주 최고가 예측 데이터

수도권_노지온주_PRICE_MAX							
	PRICE_MAX	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MAX	사과_PRICE_MAX	황금향_PRICE_MAX
count	313	313	313	313	313	313	313
mean	8507.727	4131.313	46215.15	178726.3	27036.67	75889.47	8684.723
std	9120.334	5700.7	64520.19	247977.2	20331.65	34859.52	9524.457
min	0	0	0	0	0	26812.5	0
25%	0	0	0	0	0	47625	0
50%	0	0	24.845	2069.9	33937.5	65750	4400
75%	17875	10288	103767.5	401449.6	41750	92000	15916.67
max	23680.56	17967.5	206764.3	799071.8	85000	165000	34500

[표 37] 비수도권 노지온주 최저가 예측 데이터

비수도권_노지온주_PRICE_MIN									
	PRICE_MIN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MIN	배_PRICE_MIN	사과_PRICE_MIN	포도_PRICE_MIN	황금향_PRICE_MIN
count	313.0	313.0	313.0	313.0	313.0	313.0	313.0	293.0	311.0
mean	1384.0	3899.7	46215.2	178726.3	3038.0	5595.8	2319.9	3138.2	3494.4
std	1450.7	5374.5	64520.2	247977.2	3365.3	5875.7	2287.8	4121.8	4248.0
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	0.0	0.0	24.8	2069.9	2555.6	5300.0	2846.7	0.0	0.0
75%	2615.8	9357.0	103767.5	401449.6	5416.7	10170.6	4064.7	4820.0	7643.3
max	4250.0	16748.0	206764.3	799071.8	11054.2	21481.3	7359.9	16000.0	15000.0

[표 38] 비수도권 노지온주 평균가 예측 데이터

비수도권_노지온주_PRICE_MEAN									
	PRICE_MEAN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MEAN	배_PRICE_MEAN	사과_PRICE_MEAN	포도_PRICE_MEAN	황금향_PRICE_MEAN
count	313.0	313.0	313.0	313.0	313.0	313.0	313.0	293.0	311.0
mean	3997.2	3899.7	46215.2	178726.3	12469.6	16460.6	12229.0	7103.7	5073.0
std	4191.7	5374.5	64520.2	247977.2	11968.8	16586.5	11825.9	9170.3	6000.9
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	0.0	0.0	24.8	2069.9	15350.2	19043.6	16262.5	0.0	0.0
75%	7907.5	9357.0	103767.5	401449.6	22435.5	28806.6	22072.6	11960.0	11544.1
max	11392.4	16748.0	206764.3	799071.8	40000.0	52493.0	34594.7	30981.3	16833.3

[표 39] 비수도권 노지온주 최고가 예측 데이터

비수도권_노지온주_PRICE_MAX									
	PRICE_MEAN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	단감_PRICE_MEAN	배_PRICE_MEAN	사과_PRICE_MEAN	포도_PRICE_MEAN	황금향_PRICE_MEAN
count	313.0	313.0	313.0	313.0	313.0	313.0	313.0	293.0	311.0
mean	7835.1	3899.7	46215.2	178726.3	24141.6	30194.3	30398.7	11012.7	6190.9
std	8107.1	5374.5	64520.2	247977.2	21626.7	30050.8	29658.8	13925.1	7226.3
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
50%	0.0	0.0	24.8	2069.9	34877.8	34250.0	38600.0	0.0	0.0
75%	15947.4	9357.0	103767.5	401449.6	41866.7	55944.4	55453.3	21260.0	14038.9
max	20515.8	16748.0	206764.3	799071.8	72000.0	84400.0	87687.0	49114.3	20000.0

[표 40] 수도권 하우스온주 최저가 예측 데이터

수도권_하우스온주_PRICE_MIN					
	PRICE_MIN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	포도_PRICE_MIN
count	290.0	290.0	285.0	290.0	269.0
mean	6515.0	117.6	4659.0	18182.2	2752.7
std	5722.9	144.6	5961.4	21932.8	2612.2
min	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	1000.0
50%	6183.3	69.3	2226.9	8645.5	2000.0
75%	10845.6	179.5	7355.8	29397.7	4000.0
max	25000.0	781.8	35133.3	79131.0	14900.0

[표 41] 수도권 하우스온주 평균가 예측 데이터

수도권_하우스온주_PRICE_MEAN					
	PRICE_MEAN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	포도_PRICE_MEAN
count	290.0	290.0	285.0	290.0	269.0
mean	11181.6	117.6	4659.0	18182.2	13578.8
std	8953.1	144.6	5961.4	21932.8	8163.4
min	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	7433.3
50%	13773.4	69.3	2226.9	8645.5	12266.7
75%	17802.5	179.5	7355.8	29397.7	20920.0
max	38800.0	781.8	35133.3	79131.0	30133.3

[표 42] 수도권 하우스온주 최고가 예측 데이터

수도권_하우스온주_PRICE_MAX					
	PRICE_MAX	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT	포도_PRICE_MAX
count	290.0	290.0	285.0	290.0	269.0
mean	16015.6	117.6	4659.0	18182.2	35909.8
std	12112.4	144.6	5961.4	21932.8	22693.9
min	0.0	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	15050.0
50%	20416.7	69.3	2226.9	8645.5	36000.0
75%	24397.7	179.5	7355.8	29397.7	50000.0
max	40000.0	781.8	35133.3	79131.0	85000.0

[표 43] 비수도권 하우스온주 최저가 예측 데이터

비수도권_하우스온주_PRICE_MIN				
	PRICE_MIN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT
count	288.0	288.0	283.0	288.0
mean	6765.3	101.4	4688.2	18183.8
std	5711.4	144.4	5972.3	22005.2
min	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0
50%	7126.3	36.6	2516.4	7941.2
75%	11737.6	155.0	7378.4	29658.0
max	18113.1	988.5	35133.3	79131.0

[표 44] 비수도권 하우스온주 평균가 예측 데이터

비수도권_하우스온주_PRICE_MEAN				
	PRICE_MEAN	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT
count	288.0	288.0	283.0	288.0
mean	11124.5	101.4	4688.2	18183.8
std	9111.3	144.4	5972.3	22005.2
min	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0
50%	13716.1	36.6	2516.4	7941.2
75%	18111.3	155.0	7378.4	29658.0
max	34400.0	988.5	35133.3	79131.0

[표 45] 비수도권 하우스온주 최고가 예측 데이터

비수도권_하우스온주_PRICE_MAX				
	PRICE_MAX	TRADE_TOTAL	SHIPMENT	PAST_SHIPMENT
count	288.0	288.0	283.0	288.0
mean	14477.3	101.4	4688.2	18183.8
std	11776.9	144.4	5972.3	22005.2
min	0.0	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0
50%	18478.3	36.6	2516.4	7941.2
75%	23051.3	155.0	7378.4	29658.0
max	55000.0	988.5	35133.3	79131.0

감귤 가격 예측을 위한 학습 데이터는 학습(Train):평가(Test)=70:30 비율로 구성하였으며, 데이터 정규화 작업으로는 MinMaxScaling(정규화) 작업을 진행했다. 먼저 학습 데이터에 스케일러(Scaler)를 피팅(fitting)한 후 해당 스케일러를 사용하여 학습 데이터와 평가 데이터를 변환하여 모델에 적용하였다.

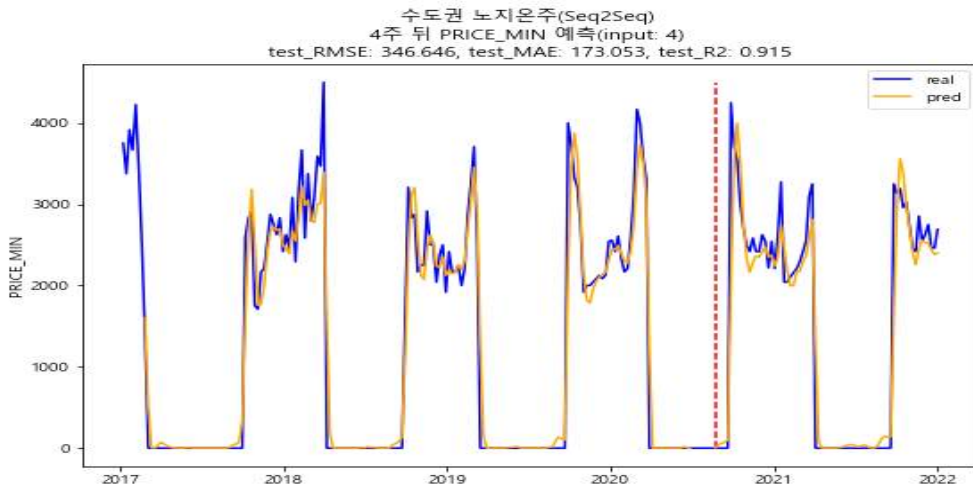
예측 모형은 '2.2. 연구방법'에서 언급했던 심층신경망 기반의 시계열 예측 모델 중 출하량 예측 시 가장 좋은 성능을 보였던 Seq2Seq 모델을 선정하여 예측력에 초점을 두어 실험을 진행했으며, 설정한 주요 변수는 [그림5]와 같다. 주요 변수 이외에도 모델 최적화를 위해 Model Capacity를 조정하는 등의 작업을 진행하였다. 비교 모델은 설명력에 초점을 둔 머신러닝 모델로 XGBoost Regressor와 Random Forest Regressor를 사용하였다.

위에서 설정된 모델을 통해 노지온주와 하우스온주에 대해 윈도우 사이즈별로 가격을 예측하는 실험을 진행한 결과는 아래와 같다. 가격 예측 모형을 구성할 때 지역 2지역(수도권, 비수도권), 예측 가격 3개(최저가, 평균가, 최고가), 시계열 모델 1개, 윈도우 사이즈는 입력 기간 4주(1개월), 출력 기간 4주(1개월), 8주(2개월), 12주(3개월)를 고려하여 총 18가지 경우에 대해 비교하는 실험을 진행하였다.

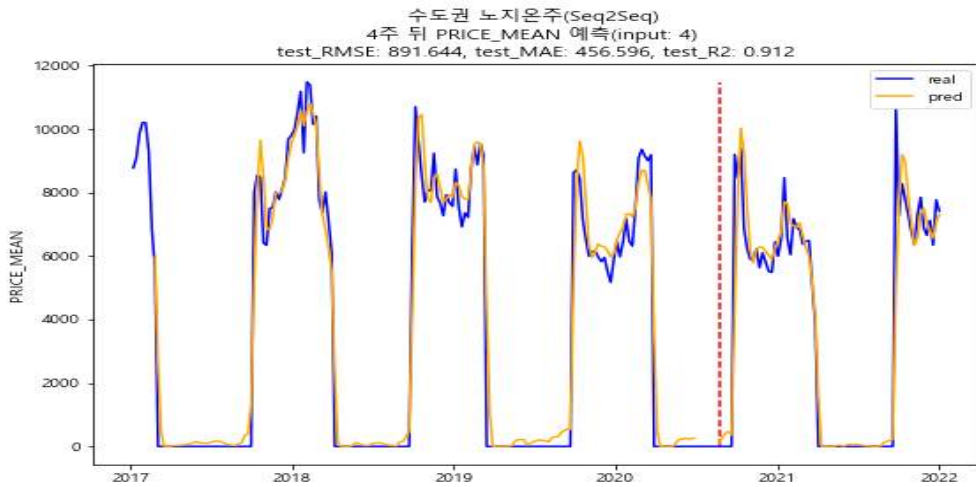
첫 번째 실험은 노주온주의 최저가, 평균가, 최고가를 예측한 결과로 [표46]~[표47], [그림45]~[그림50]과 같이 입력을 4주(1개월)로 하고 출력을 4주(1개월), 8주(2개월), 12주(3개월)로 비교하였다. 수도권과 비수도권 모두 8주와 12주 가격을 예측할 때보다 4주 가격을 예측했을 때 성능이 가장 우수하게 나타났다. 수도권과 비수도권 모두 높은 예측 성능을 보였지만 최저가와 평균가 관점에서는 비수도권, 최고가 관점에서는 수도권에 대한 예측 성능이 더 우수하게 나타났다.

[표 46] 수도권 노지온주 가격 예측 결과: 입력-4주, 출력-4주, 8주, 12주

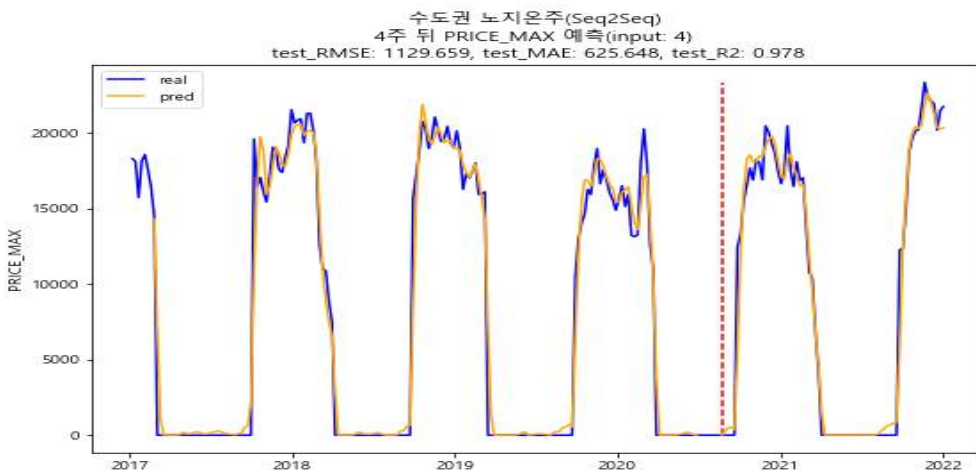
수도권_노지온주											
Input term	Output term	Model	RMSE			MAE			Adj R2-Score		
			PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX
4	4	XGBoost	501.32	1526.5	3360.47	280.11	778.75	2720.1	0.863	0.837	0.890
		Random Forest	643.93	1619.8	3197.02	35.36	745.07	2245.6	0.785	0.789	0.866
		Seq2Seq	346.64	891.64	1129.65	173.05	456.59	625.64	0.915	0.912	0.978
4	8	XGBoost	729.78	1321.24	2807.44	490.707	845.672	1930.14	0.661	0.826	0.863
		Random Forest	845.11	1573.3	2997.40	438.91	842.52	1944.2	0.628	0.801	0.891
		Seq2Seq	478.38	1199.7	1409.13	271.54	642.11	856.36	0.835	0.837	0.965
4	12	XGBoost	675.237	1352.35	2845.73	389.444	919.345	1896.37	0.706	0.812	0.876
		Random Forest	831.443	1855.96	3221.23	514.122	1045.42	2177.65	0.639	0.722	0.873
		Seq2Seq	338.498	1396.01	1343.83	188.01	793.342	828.526	0.914	0.771	0.966



[그림 45] 수도권 노지온주 최저가 예측 Best result



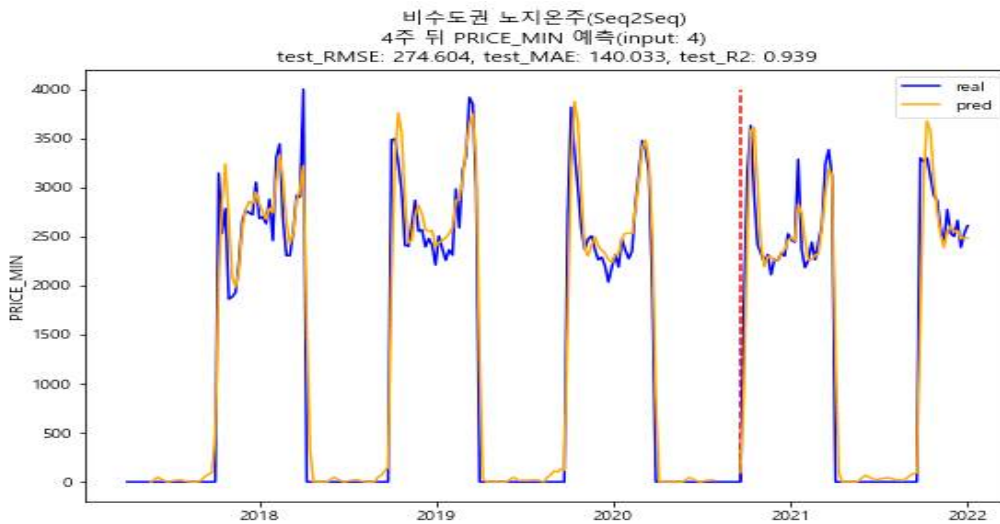
[그림 46] 수도권 노지온주 평균가 예측 Best result



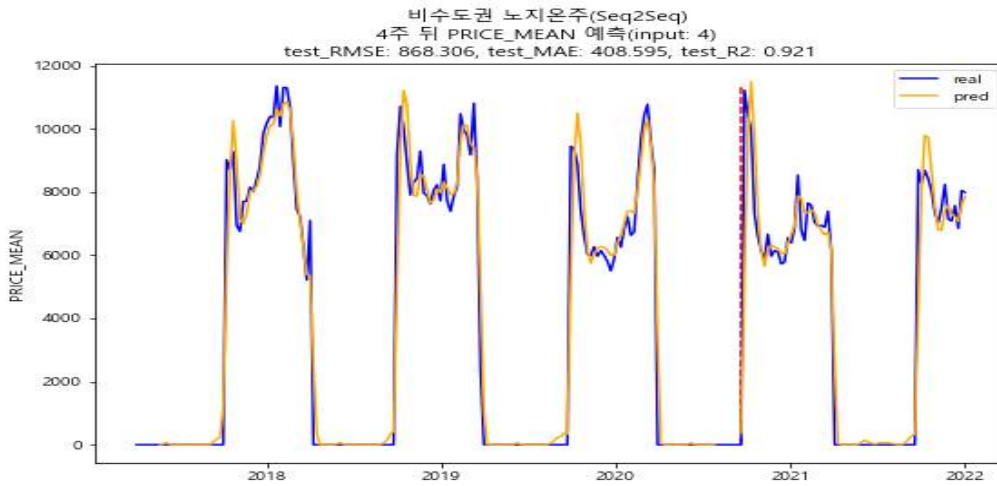
[그림 47] 수도권 노지온주 최고가 예측 Best result

[표 47] 비수도권 노지온주 가격 예측 결과: 입력-4주, 출력-4주, 8주, 12주

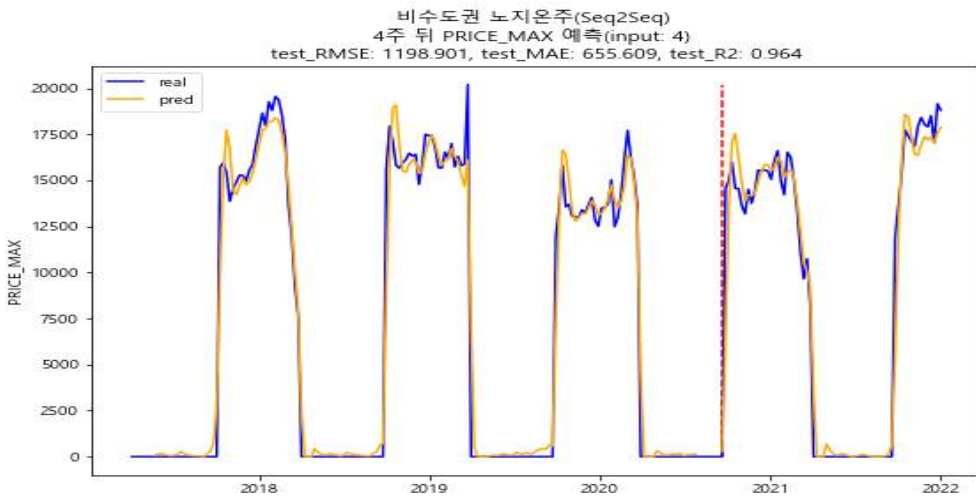
비수도권 노지온주											
Input term	Output term	Model	RMSE			MAE			Adj R2-Score		
			PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX
4	4	XGBoost	562.712	1482.18	2660.44	319.377	765.721	1586.44	0.893	0.862	0.845
		Random Forest	681.445	1509.79	2681.02	422.697	772.487	1444.10	0.681	0.791	0.852
		Seq2Seq	274.60	868.30	1198.90	140.03	408.59	655.60	0.939	0.921	0.964
4	8	XGBoost	638.12	1630.30	2391.7	317.57	854.589	1410.97	0.774	0.811	0.905
		Random Forest	772.6	1615.53	2448.80	531.345	848.161	1361.64	0.59	0.756	0.873
		Seq2Seq	435.7	1195.99	1634.8	241.17	686.14	1017.41	0.842	0.842	0.93
4	12	XGBoost	729.021	1630.32	2493.00	388.12	864.516	1366.24	0.815	0.748	0.866
		Random Forest	575.682	1740.81	3033.90	309.947	924.432	1740.15	0.636	0.784	0.846
		Seq2Seq	366.23	929.69	1828.33	207.20	532.61	991.58	0.886	0.9	0.909



[그림 48] 비수도권 노지온주 최저가 예측 Best result



[그림 49] 비수도권 노지온주 평균가 예측 Best result

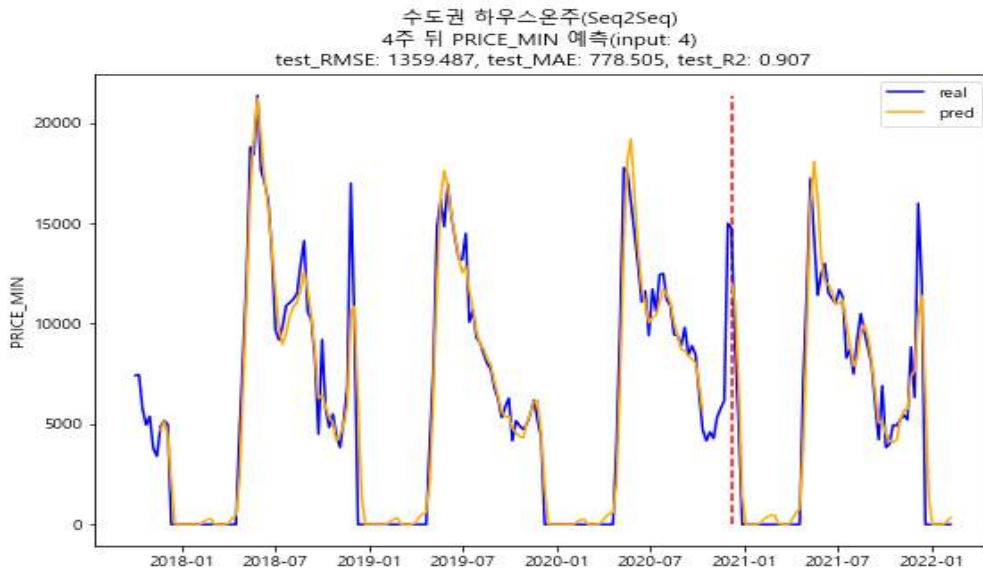


[그림 50] 비수도권 노지온주 최고가 예측 Best result

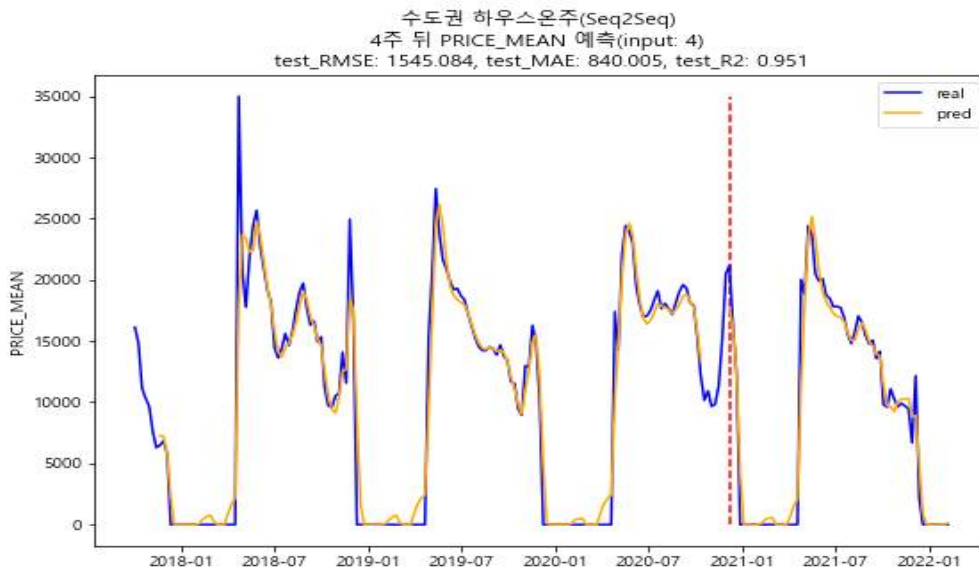
두 번째 실험은 하우스온주의 최저가, 평균가, 최고가를 예측한 결과로 [표 48]~[표49], [그림51]~[그림56]과 같이 입력을 4주(1개월)로 하고 출력을 4주(1개월), 8주(2개월), 12주(3개월)로 비교하였다. 수도권과 비수도권 모두 8주와 12주 가격을 예측할 때보다 4주 가격을 예측했을 때 성능이 가장 우수하게 나타났다. 수도권/비수도권 모두 높은 예측 성능을 보이나 최저가 관점에서는 비수도권, 평균가와 최고가 관점에서는 수도권에 대한 예측 성능이 더 우수하게 나타났다.

[표 48] 수도권 하우스온주 가격 예측 결과: 입력-4주, 출력-4주, 8주, 12주

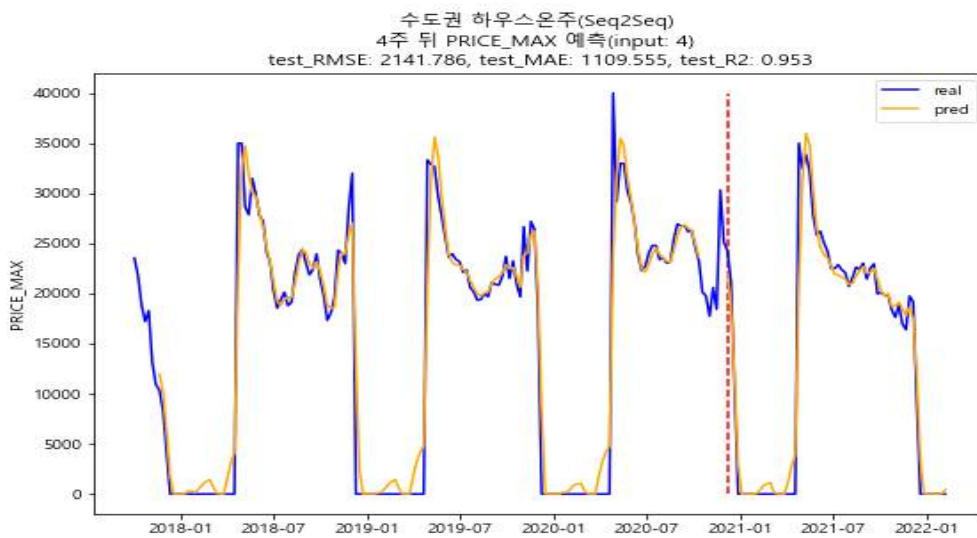
수도권_하우스온주											
Input term	Output term	Model	RMSE			MAE			Adj R2-Score		
			PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX
4	4	XGBoost	3009.29	3208.53	3504.15	1895.16	2040.54	2179.42	0.795	0.878	0.860
		Random Forest	2906.00	3726.22	5295.97	1638.00	2055.53	2786.10	0.679	0.786	0.788
		Seq2Seq	1359.48	1545.08	2141.78	778.505	840.005	1109.55	0.907	0.951	0.953
4	8	XGBoost	3050.8	3673.7	3930.99	1987.1	2275.8	2471.1	0.606	0.786	0.867
		Random Forest	3618.3	3726.2	5295.97	1977.6	2055.5	2786.1	0.524	0.788	0.783
		Seq2Seq	2079.2	2260.1	2575.07	1281.3	1295.2	1343.8	0.772	0.891	0.93
4	12	XGBoost	3009.9	4327.59	4657.16	1980.97	2561.30	2788.97	0.602	0.669	0.799
		Random Forest	3626.9	4487.19	6270.88	1985.56	2413.50	3047.26	0.521	0.703	0.712
		Seq2Seq	1904.2	2039.93	3676.24	1219.52	1204.43	2331.05	0.798	0.908	0.85



[그림 51] 수도권 하우스온주 최저가 예측 Best result



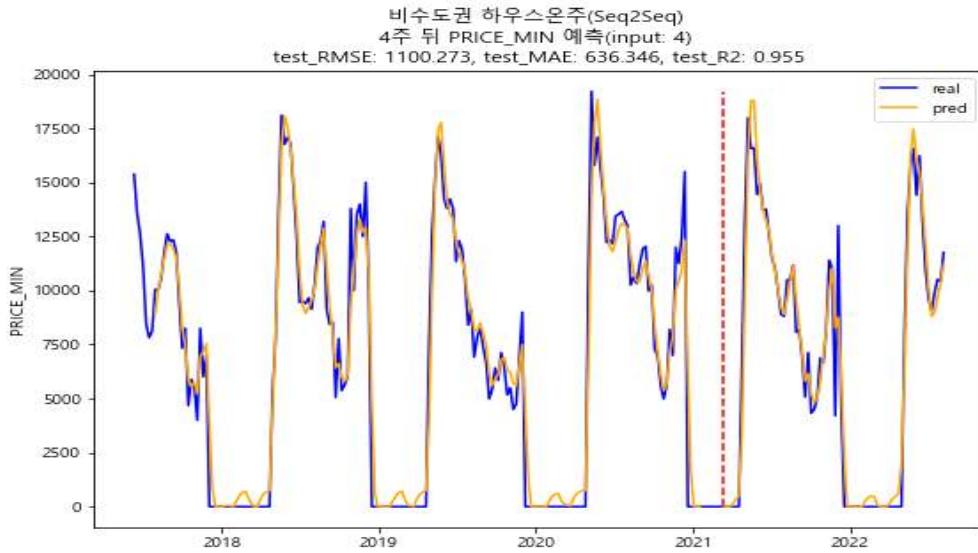
[그림 52] 수도권 하우스온주 평균가 예측 Best result



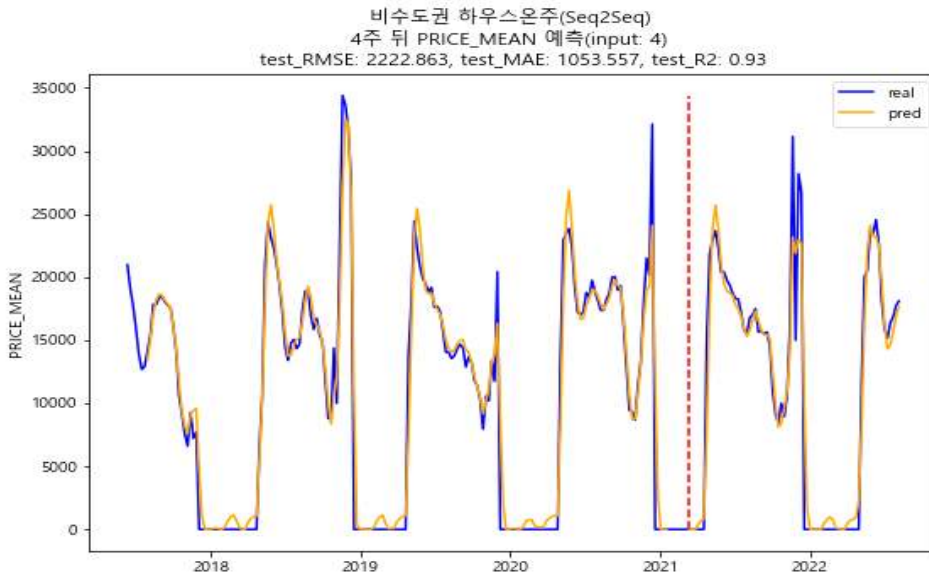
[그림 53] 수도권 하우스온주 최고가 예측 Best result

[표 49] 비수도권 하우스온주 가격 예측 결과: 입력-4주, 출력-4주, 8주, 12주

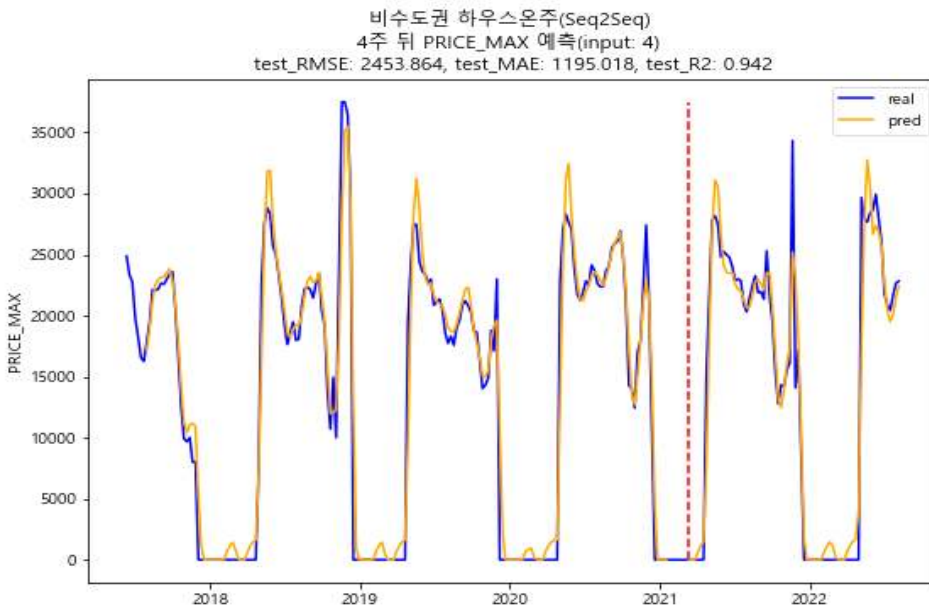
비수도권_하우스온주											
Input term	Output term	Model	RMSE			MAE			Adj R2-Score		
			PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX	PRICE_MIN	PRICE_MEAN	PRICE_MAX
4	4	XGBoost	2049.41	3777.75	3491.03	1272.96	2004.94	2012.42	0.842	0.795	0.856
		Random Forest	2265.36	4349.85	4230.54	1283.44	2065.95	1916.47	0.853	0.792	0.868
		Seq2Seq	1100.27	2222.86	2453.86	636.346	1053.55	1195.01	0.955	0.93	0.942
4	8	XGBoost	2269.47	4067.54	4050.46	1454.85	2338.67	2457.98	0.835	0.779	0.868
		Random Forest	2609.11	4222.18	4334.60	1412.62	2225.43	2268.48	0.805	0.803	0.861
		Seq2Seq	1597.66	3063.10	3280.56	984.368	1799.74	1774.21	0.906	0.866	0.895
4	12	XGBoost	2430.01	4440.10	4593.30	1542.11	2456.31	2801.89	0.816	0.722	0.874
		Random Forest	2698.17	5306.11	4323.41	1452.99	2511.07	2163.82	0.793	0.689	0.817
		Seq2Seq	1485.55	3059.58	3002.83	832.146	1820.80	1669.94	0.916	0.863	0.91



[그림 54] 비수도권 하우스온주 최저가 예측 Best result



[그림 55] 비수도권 하우스온주 평균가 예측 Best result

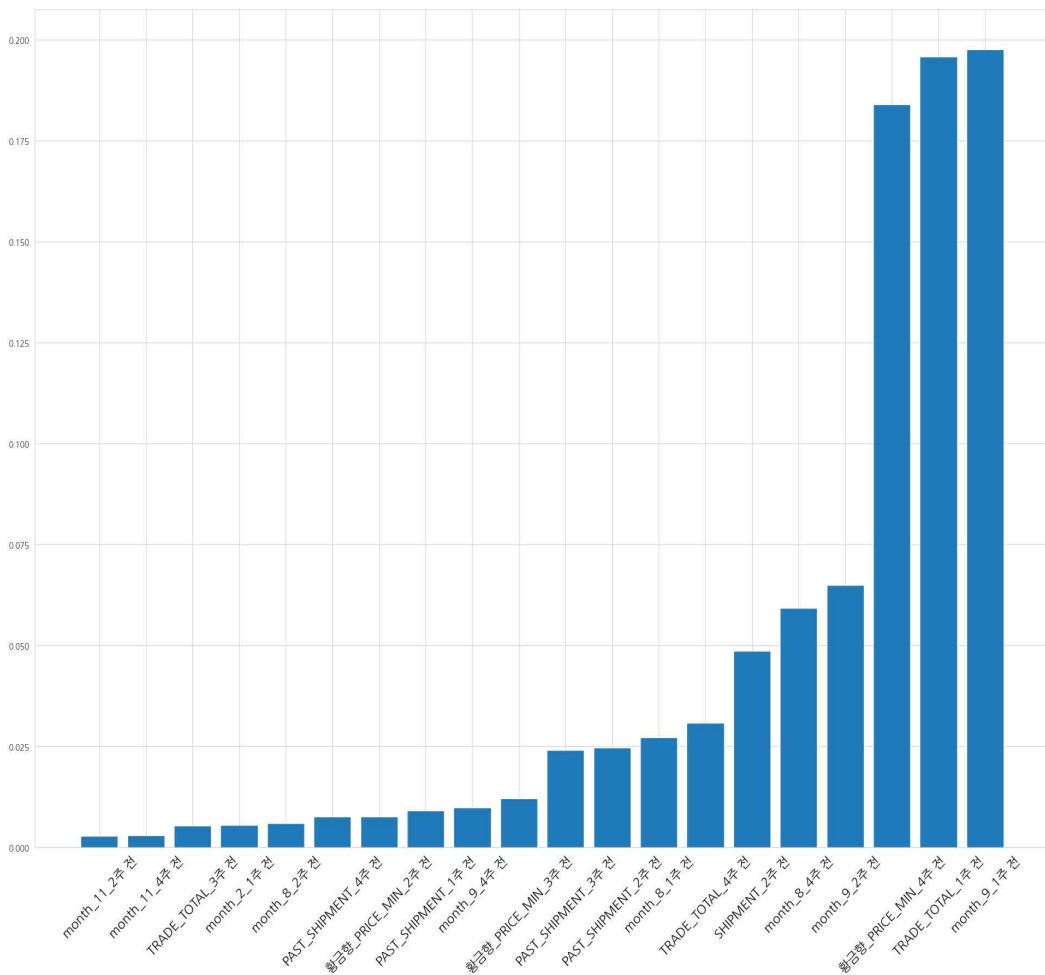


[그림 56] 비수도권 하우스온주 최고가 예측 Best result

이를 통해 본 연구에서는 감귤의 가격 예측 시 설명력에 초점을 둔 머신러닝 모형보다 예측력 기반의 심층신경망 모형의 성능이 좋게 나타나는 것을 확인할 수 있었다.

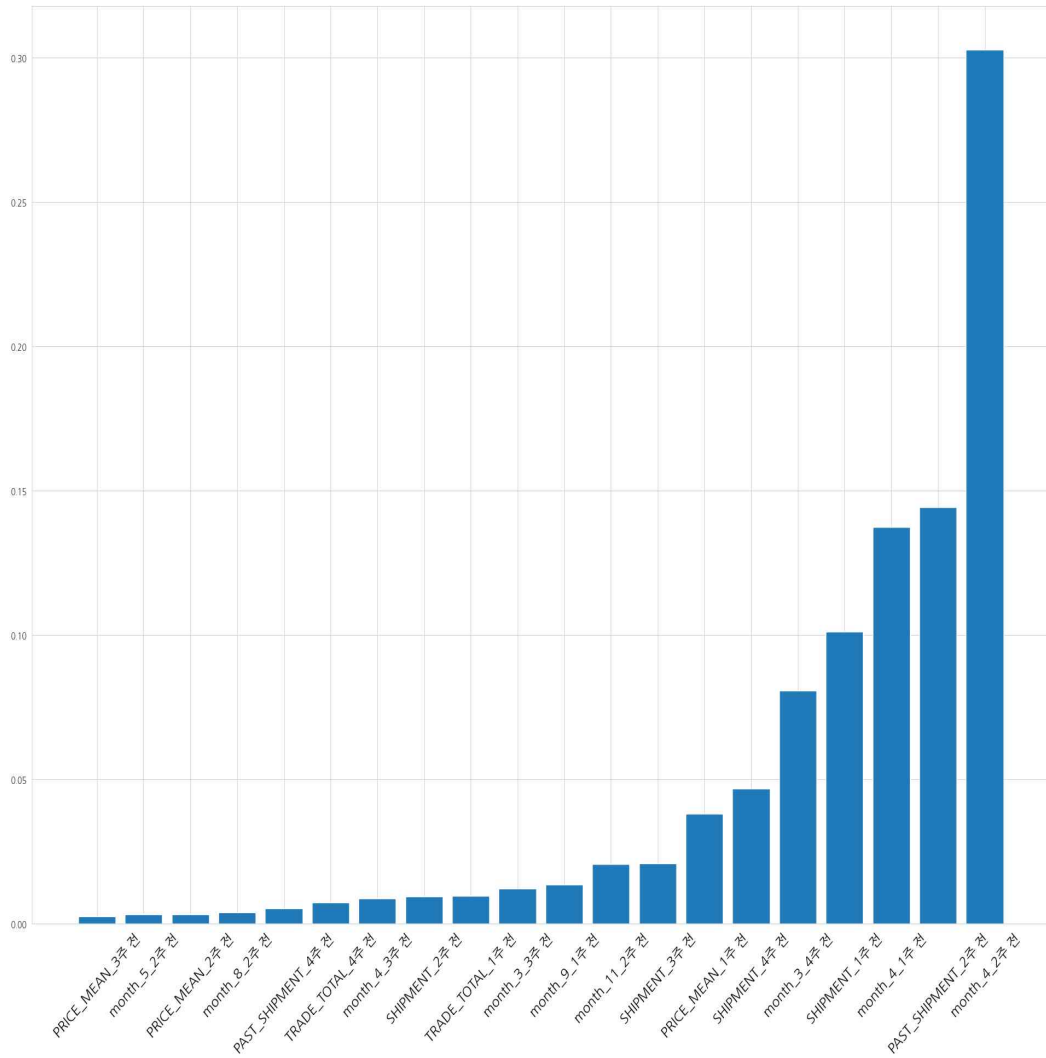
다음으로 설명력 기반의 모형을 통해 변수 중요도(Feature Importance)를 알아 보았다. 감귤 가격 예측 시 설명력 기반의 모형은 Random Forest보다 XGBoost가 성능이 우수하게 나타났으며, 노지온주는 비수도권 지역에 입력 기간을 4주(1개월), 출력 기간을 4주(1개월)로 설정하여 최소가격을 예측했을 때 Adjusted R2-Score가 가장 높게 나타났고, 하우스온주는 수도권 지역에 입력 기간을 4주(1개월), 출력 기간을 4주(1개월)로 설정하여 평균가격을 예측했을 때 Adjusted R2-Score가 가장 높게 나타났다.

노지온주 가격 예측 시 가장 우수한 성능을 보인 모형의 변수 중요도를 15위까지 출력해 보면 [그림57]과 같이 거래 시기 월 정보가 가장 큰 영향력을 보였으며 그 뒤로는 총거래량, 경쟁 과일(황금향) 판매가격, 누적 출하량 순으로 변수 중요도가 높게 나타났다.



[그림 57] 머신러닝 기반 노지온주 가격 예측 Best Model의 변수 중요도

하우스온주 가격 예측 시 가장 우수한 성능을 보인 모형의 변수 중요도를 15 위까지 출력해 보면 [그림58]과 같이 거래 시기 월 정보가 가장 큰 영향력을 보였으며 그 뒤로는 출하량, 누적 출하량, 평균가격, 총거래량 순으로 변수 중요도가 높게 나타났다.



[그림 58] 머신러닝 기반 하우스온주 가격 예측 Best Model의 변수 중요도

5. 결론

5.1. 요약

한국의 농산물 시장은 국가에서 관리하는 공영도매시장을 중심으로 경제적인 중요성을 가지고 있다. 특히 지역마다 특산물을 보유하고 있으며, 각 지역의 특산물은 지역경제에 큰 영향을 미치고 있다. 제주에서는 감귤 생산, 공급의 안정화는 생산자 이외에도 유통업자, 소비자, 정부, 학계 등 다양한 이해 관계자에게 중요한 주제이다.

기존 농산물에 관한 연구 현황을 살펴보면 전반적인 농작물의 생산이나 다른 과육의 생산성 관련 연구논문은 다수 있으나 제주 밀감의 대부분을 차지하는 노지온주와 하우스온주에 대해 인공지능을 활용하여 생산량이나 가격을 예측하는 분석 연구는 부재하다. (강민철, 이승호. 2019; 김세원, 김영희. 2021; 정대호, 조영열. 2022; 박종현 등. 2023) 이에 본 연구에서는 기존 농산물 생산과 가격 예측 연구의 한계점을 개선하여 제주 감귤 출하와 가격 변동 특성에 맞춘 감귤 출하량 및 가격 예측 모델을 개발하였다.

감귤 출하량 및 가격 예측에는 다양한 요인 변수가 활용되며 시장의 변동성, 재배 환경(기온, 습도, 일사량, 강우량, 적설량 등), 경쟁 과일의 수요 및 공급 등 다양한 요인 등이 해당한다. 본 연구에서는 과일 출하량 및 가격 변화와 상관관계가 있는 다양한 국·내외 영향요인을 분석하고, 시간에 따른 각 요인을 독립변수로 하여 출하량과 가격을 예측하는 모형으로 설계했다.

본 연구에서는 예측 정확도 향상을 위해 심층학습 기반의 LSTM, Seq2Seq, Attention 모델들을 활용하여 감귤 출하량 및 가격 예측을 수행하고, 노지온주와 하우스온주 품종에 대해 윈도우 사이즈에 따른 출하량과 가격의 예측 성능을 비교하는 방법으로 실험을 진행하였다.

출하량 예측 모델을 시뮬레이션해본 결과, 노지온주의 경우 예측 모델은 Seq2Seq 모형을 사용하여 입력을 24주(6개월)로 하고 출력을 4주(1개월)로 설정하였을 때 예측 성능이 가장 우수하게 나타났다. 반면에, 하우스온주의 경우 예측 모델은 Seq2Seq 모형을 사용하여 입력을 20주(5개월)로 하고 출력을 4주(1개

월)로 설정하였을 때 예측 성능이 가장 우수하게 나타났다.

가격 예측 모델을 시뮬레이션해본 결과, 수도권과 비수도권 모두 8주와 12주 가격을 예측할 때보다 4주 가격을 예측했을 때 성능이 가장 우수하게 나타났다. 노지온주의 경우 수도권과 비수도권 모두 높은 예측 성능을 보였지만 최저가와 평균가 관점에서는 비수도권, 최고가 관점에서는 수도권에 대한 예측 성능이 더 우수하게 나타났다. 반면에, 하우스온주의 경우는 최저가 관점에서는 비수도권, 평균가와 최고가 관점에서는 수도권에 대한 예측 성능이 더 우수하게 나타났다.

또한, 출하량 예측 모형과 가격 예측 모형 모두 설명력에 초점을 둔 머신러닝 모형보다 예측력 기반의 심층신경망 모형의 성능이 좋게 나타났으며, Random Forest Regressor 모형보다 XGBoost Regressor 모형의 성능이 더욱 좋게 나타났다. 변수 중요도는 실험 모형별로 모두 다르게 나타났으나, 대부분 출하 시기와 거래 시기 변수의 중요도가 높게 나타났다.

5.2. 연구의 한계 및 향후 과제

첫 번째는, 본 연구에서는 노지온주과와 하우스온주의 감귤에 대한 출하량 및 가격 예측 모델을 개발하고 시뮬레이션을 수행하였으나, 다른 품종의 감귤에 대해서는 진행하지 않았다. 향후 연구에서는 다른 감귤 품종에 대해서도 출하량 및 가격 예측 모델을 구축하여 비교 분석하는 것이 필요하다. 다양한 감귤 품종의 특성과 재배 환경을 고려하여 예측 모델을 개발함으로써 특정 감귤 품종이 아닌 전체 감귤 시장의 전반적인 예측이 가능할 것이다.

두 번째는, 본 연구에서는 출하량과 가격 예측 시 제주 전체를 대상으로 분석하였지만, 지역을 더 세분화하여 행정동 또는 읍면동 단위로 예측 모형을 고도화할 수 있을 것이다. 지역별로 감귤의 생산 조건, 시장 특성, 수요 및 공급의 차이 등이 다를 수 있으므로 지역별 예측 모델을 개발하여 더욱 정확한 예측을 수행할 수 있을 것이다.

세 번째는 기존에 연구가 많이 되었던 시계열 모델이 아닌 최근 연구되고 있는 모델을 활용하여 비교해보는 것이다. 김동건, 김광수. (2021)에 따르면 최근에는 트랜스포머 모델이 시계열 예측 모델로서 주로 연구되고 있으며, 기존 트랜스

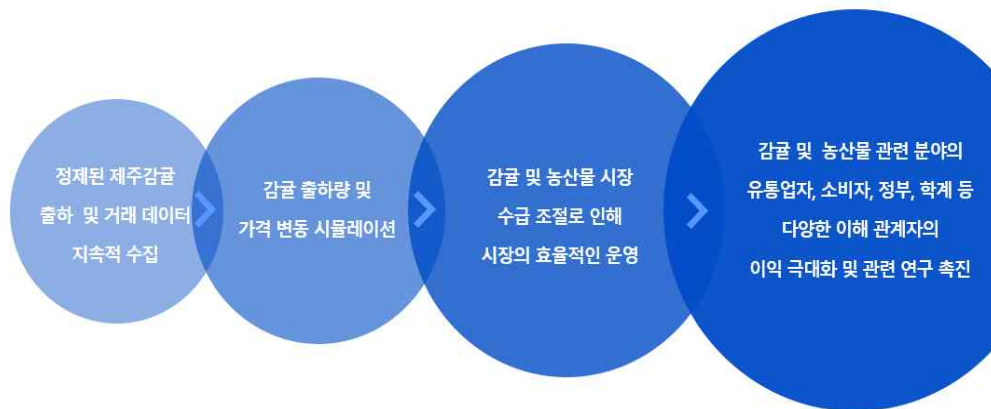
포머 모델의 예측 순차를 출력할 때 출력 결과를 다시 입력하는 자가회귀 구조로 되어 있다는 한계점을 개선하여 더 정확한 시계열 예측을 위해 스타일 변환 기법에 착안한 순차 디코딩 모델을 제안하였다. 이 구조는 자가회귀 방식의 기존의 트랜스포머의 디코더 구조와 다르게, 예측 순차를 한꺼번에 출력하기 때문에 더 먼 시점의 정보를 좀 더 정확히 예측할 수 있다는 장점이 있다.

향후 연구에서는 위와 같은 한계를 보완하고 개선하여 더 정확하고 신뢰할 수 있는 감귤 출하량 및 가격 예측 모델의 개발과 응용을 위해 노력할 것이다.

5.3. 기대 효과

본 연구의 결과는 감귤 출하량 및 가격 예측에 영향을 미치는 요인 변수의 중요성을 확인하고, 다양한 시계열 모델의 윈도우 사이즈를 조절하여 시나리오별 감귤의 출하량과 가격의 시뮬레이션이 가능하다. 심층학습 기반으로 출하량 및 가격 예측 모델들의 결과를 시뮬레이션하고 비교 제시함으로써, 농산물 시장의 효율적인 운영과 농산물 생산자 및 소비자의 이익 극대화 관련 연구 및 실무 분야에 도움이 될 것으로 기대된다.

학문, 경제, 사회 전반 긍정적인 파급력 기대



[그림 59] 감귤 출하량 및 가격 예측 모형의 파급 효과

6. 참고문헌

- [1] 농촌진흥청. (2022). 고품질 감귤 생산지침(농업기술길잡이189), 휴먼컬처아리량, 2022.04.08.
- [2] 강민철, 이승호. (2019). 제주도에서 생육시기별 기후변동이 감귤 생산량에 미치는 영향. 기후연구 제14권, 1호, pp. 15-16.
- [3] 김동건, 김광수. (2021). 시계열 예측을 위한 스타일 기반 트랜스포머. KTSDE 2021, vol.10, no.12, 통권 111호, pp. 579-586.
- [4] 김세원, 김영희. (2021). 작물 생산량 예측을 위한 머신러닝기법 활용 연구. 한국산학기술학회논문지 제22권 제7호.
- [6] 박종현, 임영우, 임도현, 최윤성, 안현철. (2023). 중장기 농산물 가격 예측을 위한 다단계 시계열 예측 모델. 한국컴퓨터정보학회 논문지제28권 제2호(통권 제227호).
- [7] 유동완, 박종범. (2022). LSTM 모델을 이용한 농산물 가격 예측에 관한 연구. 한국정보통신학회 2022년도 춘계종합학술대회 논문집 제26권 제1호.
- [8] 정대호, 조영열. (2022). LSTM 네트워크를 이용한 한라봉가격 예측 모델. 원예과학기술지 제40권 제5호.
- [9] 정소영, 양승룡. (2022). 농산물 도매시장 효율성 국제 비교분석. 한국농식품정책학회 학술대회 논문집, pp. 507-540.
- [10] 최가은. (2022). 농산물 가격 예측에서 오차를 줄이기 위한 요소 탐색. 한국정보과학회 2022 한국소프트웨어종합학술대회 논문집, pp. 1,540-1,542.
- [11] Bae, K. T. and Kim, C. J. (2016). An Agricultural Estimate Price Model of Artificial Neural Network by Optimizing Hidden Layer. Journal of the Korean Society of Information Technology, 14(12).
<http://doi.org/10.14801/jkiit.2016.14.12.161>
- [12] Bai, S., Kolter, J. Z. and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint, arXiv:1803.01271, <https://doi.org/10.48550/arXiv.1803.01271>.
- [13] Deshmukh, A. A., Srivatsa, A., Ashwitha, A., Monteiro, A., and Gajakosh, C.

- (2022). Crop Yield Prediction to Achieve Precision Agriculture using Machine Learning. In 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), pp. 1-6.
- [14] Hochreiter, S. and Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. *Advances in Neural Information Processing Systems 9 (NIPS 1996)*.
- [15] Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9, pp. 1735-1780.
<https://doi.org/10.1162/neco.1997.9.8.1735>.
- [16] Jang, S., Chun, H., Cho, I. and Kim, D. (2017). A study on cabbage wholesale price forecasting model using unstructured agricultural meteorological data. *Journal of the Korean Data and Information Science Society*, 28(3), pp. 617-24. <https://doi.org/10.7465/jkdi.2017.28.3.617>.
- [17] Jung, D. H., and Cho, Y. Y. (2022). A Prediction Model for Hallabong Tanager Product Prices using LSTM (Long Short-term Memory) Network. *Horticultural Science and Technology*, 40(5), pp 571-577.
<http://doi.org/10.7235/HORT.20220051>
- [18] Kim, J. and Lee, J. I. (2019). A Study on Agricultural Price Prediction System based on Deep Learning. *The Journal of Korean Institute of Information Technology*, 17(6), pp. 27-34.
<http://doi.org/10.14801/jkiit.2019.17.6.27>
- [19] Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint, arXiv:1412.6980.
- [20] Kuriakose, S. M., and Singh, T. (2022). Indian Crop Yield Prediction using LSTM Deep Learning Networks. 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT).
<http://doi.org/10.1109/ICCCNT54827.2022.9984407>
- [21] Lee, Y. S., Kim, J. J., Kim, G. S., An, D. H. and Jo, J. S. (2013). Influence on aggregate general price level of agricultural pricing and policy

- implications. Policy Research Report, Korea Rural Economic Institute (KREI).
<http://repository.krei.re.kr/handle/2018.oak/20472>
- [22] Liao, S., Wang, Z., Luo, Y. and Liang, H. (2021). Locational marginal price forecasting using Transformer-based deep learning network. 2021 40th Chinese Control Conference (CCC).
<https://doi.org/10.23919/CCC52363.2021.9549619>.
- [23] Olah, C. (2015). Understanding LSTM networks. Github, 2015.
<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [24] Olah, C. and Carter, S. (2016), Attention and Augmented Recurrent Neural Networks. Distill, 2016.
<http://doi.org/10.23915/distill.00001>
<https://distill.pub/2016/augmented-rnns/#attentional-interfaces>.
- [25] Purohit, S. K., Panigrahi, S., Sethy, P. K. and Behera, S. K. (2021). Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. Applied Artificial Intelligence 2021, pp. 1388-1406.
<https://doi.org/10.1080/08839514.2021.1981659>
- [26] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G. and Cottrell, G. (2017).c A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. International Joint Conference on Artificial Intelligence (IJCAI 2017). <https://doi.org/10.48550/arXiv.1704.02971>.
- [27] Sahoo, D., Sood, N., Rani, U., Abraham, G., Dutt, V., and Dileep, A. D. (2020). Comparative analysis of multi-step time-series forecasting for network load dataset. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-7.
- [28] Sahoo, D., Sood, N., Rani, U., Abraham, G., Dutt, V., and Dileep, A. D. (2020). Comparative analysis of multi-step time-series forecasting for network load dataset. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp.

- 1-7. <http://doi.org/10.1109/ICCCNT49239.2020.9225449>.
- [29] Shin, S., Lee, M. and Song, S. (2018). A Prediction Model for Agricultural Products Price with LSTM Network. Journal of Korean Contents Society, 18(11). <https://doi.org/10.5392/JKCA.2018.18.11.416>.
- [30] Sutskever, I., Vinyals, O., and Quoc, V. L. (2014). Sequence to Sequence Learning with Neural Networks. Advances in Neural Information Processing Systems(NIPS), 27.
- [31] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
- [32] 관세청. (2022). 관세청_품목별 수출입실적(GW). 공공데이터포털. 2022.07.15. <https://www.data.go.kr/data/15101609/openapi.do>.
- [33] 기상청. (2021). 기상청_지상(중관, ASOS) 일자료 조회서비스. 공공데이터포털. 2021.07.28. <https://www.data.go.kr/tcs/dss/selectApiDataDetailView.do?publicDataPk=15059093>.
- [34] 제주감귤출하연합회. (2022). 9대도매시장 법인별 현황. 감귤출하신고시스템. <http://jcitrus.or.kr/shipment/9major.htm>.
- [35] 제주감귤출하연합회. (2022). 감귤출하상황. 감귤출하신고시스템. <http://citrus2.or.kr/ship/print01.php>.
- [36] 제주특별자치도 친환경농업정책과. (2021). 2021년 제주농축산식품현황. 제주특별자치도. 2021.08.24. <https://www.jeju.go.kr/news/news/data.htm?act=view&seq=1326665>.
- [37] 통계청 농어업통계과. (2023). 2022년 농작물생산조사. KOSIS국가통계포털. <https://www.narastat.kr/metasvc/index.do?confmNo=114004&inputYear=2023>. 2023.5.20.
- [38] 한국농수산물유통공사. (2023). 정산가격 기간별 도매시장별 품목별 총물량총금액. aT 도매시장 통합홈페이지. <https://at.agromarket.kr/openApi/apiInfoDtl.do?apiSeq=24>.

[39] 한국천문원. (2022). 한국천문연구원_특일 정보. 공공데이터포털. 2022.03.31.
<https://www.data.go.kr/data/15012690/openapi.do>.

Development of a Predictive Model for Citrus Shipments and Prices, and Analysis of Influencing Factors

Kim Seong Yul

Department of Computational Statistics
The Graduate School
Jeju National University

Abstract

The primary industry scale in Jeju island is 4.2249 trillion won, with agricultural incomes accounting for 1.7 trillion won (40.3%). Despite the unique characteristics of being surrounded by the sea, agriculture has a high proportion in Jeju, and citrus fruits account for over half of the agricultural imports, generating 950.8 billion won (55.92%) in revenue. Therefore, the stabilization of citrus production and supply in Jeju is an important topic not only for producers but also for various stakeholders such as distributors, consumers, government, and academia. While there are numerous research papers on overall crop production or productivity related to other fruits, there is a lack of analytical studies using artificial intelligence to predict shipments

and prices for open-field and greenhouse-grown citrus, which represent the majority of Jeju's citrus production. In this study, we aim to develop shipments and price prediction models for open-field and greenhouse-grown citrus by referring to techniques from crop production and price prediction research papers and tailoring them to Jeju's citrus. Various factor variables are utilized for citrus shipments and price prediction, including market volatility, cultivation environment (temperature, humidity, sunshine duration, precipitation, snowfall, etc.), demand and supply of competing fruits, and other factors. In this study, we analyze various domestic and international factors that are correlated with changes in fruit shipments and prices and design a model that predicts the shipments and prices by using each factor as an independent variable over time. Domestic factors used in this study include citrus shipments, price status, weather information, and holidays, while international factors include imported fruit data of competing fruits. To select the optimal prediction model, various time series models were compared and analyzed. In order to improve the prediction accuracy, deep learning-based LSTM, Seq2Seq, and Attention-based models were utilized to perform citrus shipments and price prediction, and the prediction accuracy and stability were evaluated. The results of this study are expected to confirm the importance of factor variables that influence citrus shipments and price prediction and enable the simulation of shipments and price scenarios by adjusting the window size of various time series models. By comparing and presenting the results of shipments and price prediction models based on deep learning, this study is expected to contribute to research and practical fields related to the efficient operation of the agricultural market and maximizing the benefits of agricultural producers and consumers.