

MindLab 과일 수요량 예측

성득명 김태훈 이호 박선영

목차

① 프로젝트 개요

1. 주제선정 및 배경
2. 개발 환경 및 기간

② 데이터 수집 및 전처리

1. 데이터 수집
2. 데이터 전처리

③ 모델 설명 및 평가

- 주요 모델 설명
1. XGBoost
 2. LightGBM
 3. Random Forest
 4. LSTM
 5. 평가

④ 웹 구현

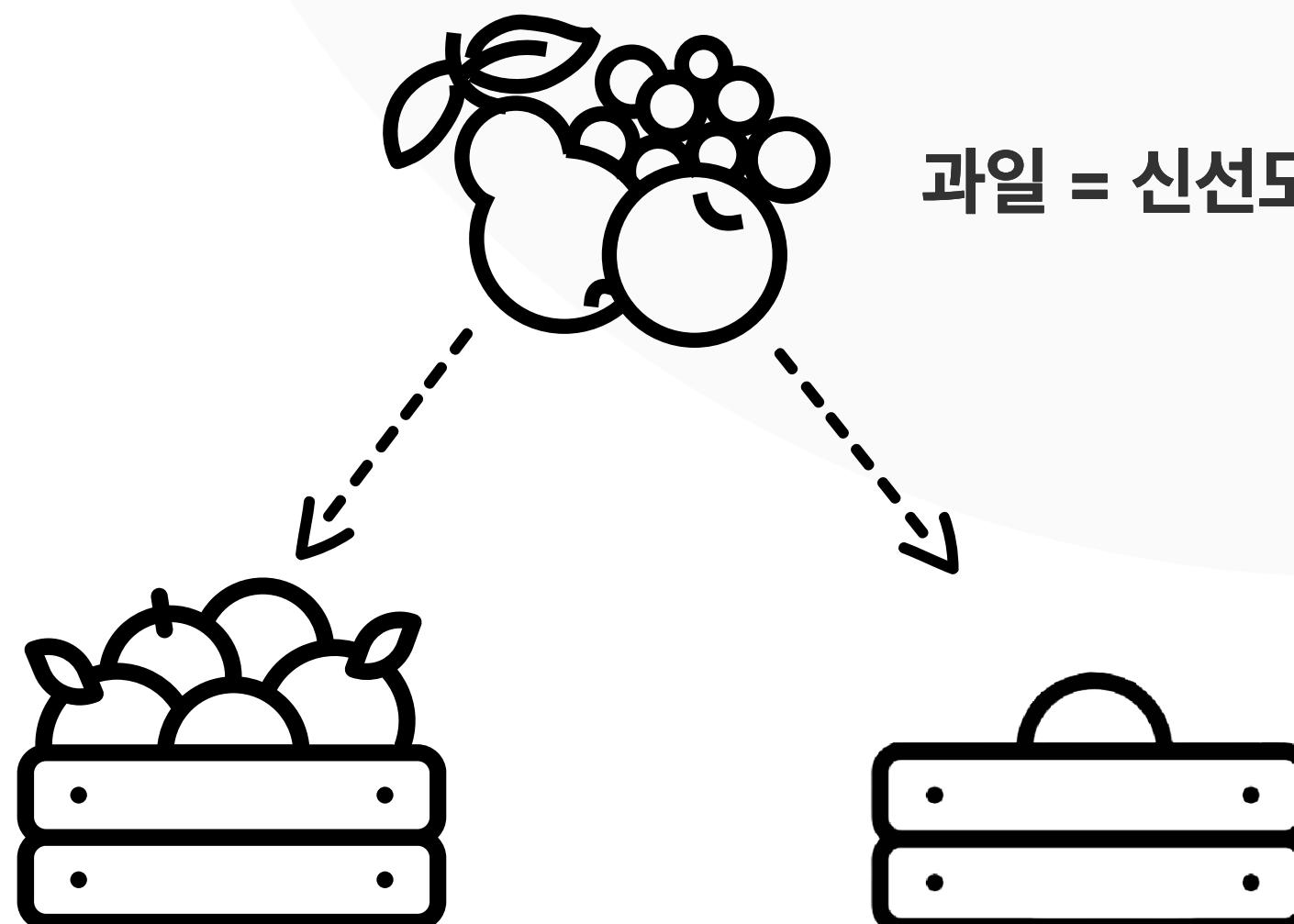
1. Streamlit

⑤ 결론 및 평가

1. 기대효과

1. 프로젝트 개요

1. 주제선정 및 배경



과일 = 신선도로 인해 재고 관리가 어려움

=

정확한 수요 예측 필요

재고 多
폐기 비용 증가(악성재고)

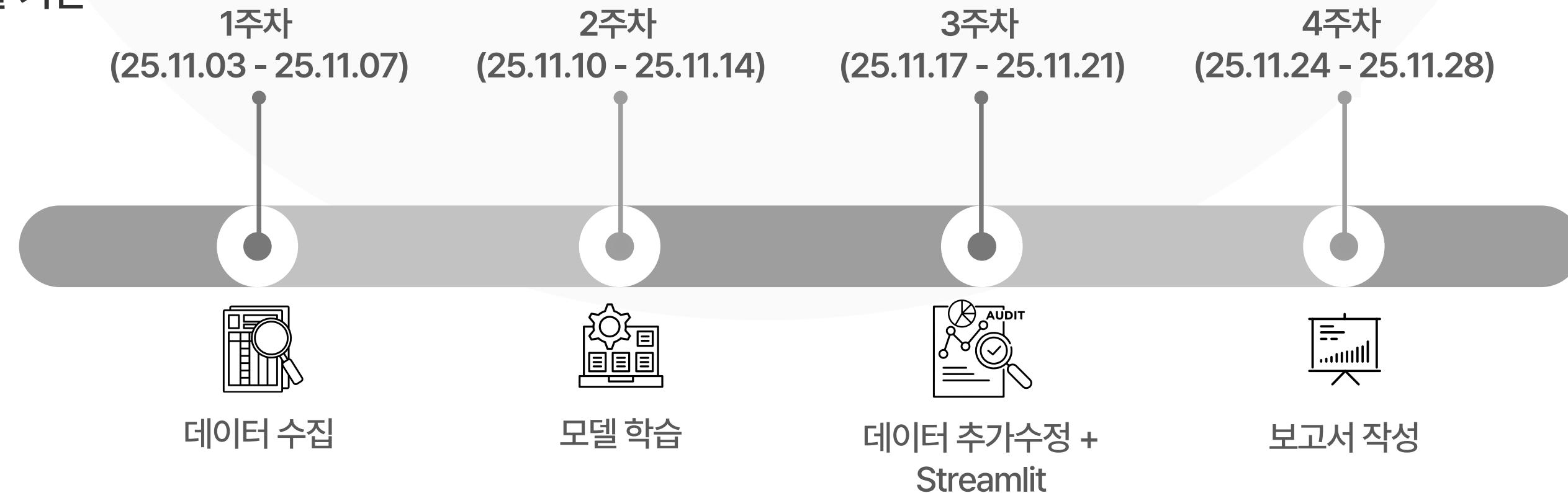
재고 小
판매 손실

학습 중인 AI 모델을
실생활에 적용해보고자 프로젝트 기획

1. 프로젝트 개요

2. 개발 환경 및 기간

- 개발 기간



- 개발 환경

GOOGLE COLAB

STREAMLIT

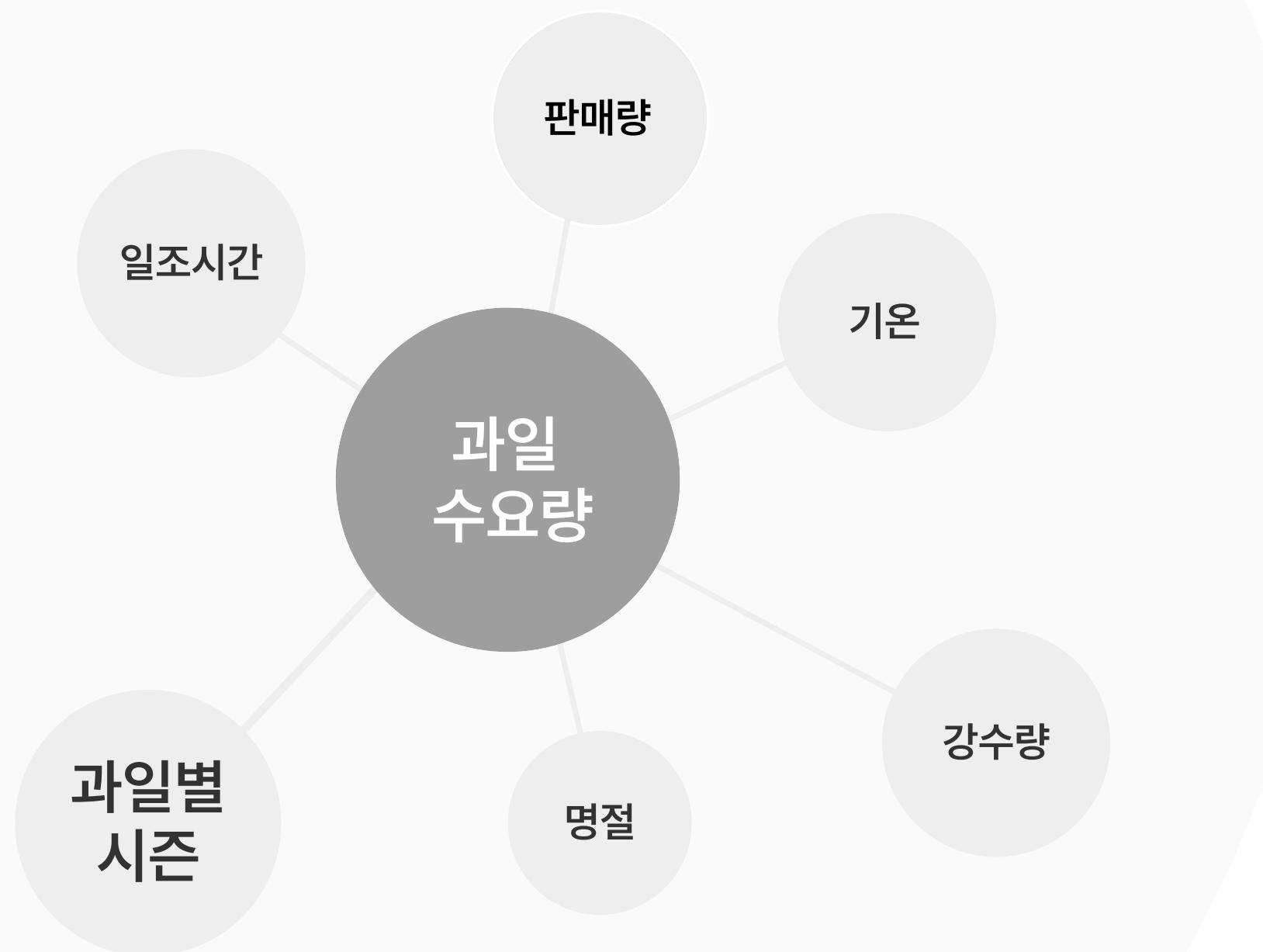
FIGMA

HUGGING FACE

GITHUB

2. 데이터 수집 및 전처리

1. 데이터 수집



데이터 준비

1. 기본 데이터: KAMIS 농수산물 도소매 가격 (판매량(kg))
2. 일기 데이터: 기후통계분석(기온, 강수량, 일조시간)
3. 수요패턴: 각 과일의 월별 패턴, 명절에 따른 수요량

- [출처]
<공공 데이터 포털>
 - 한국 농수산 식품유통 공사, 한국농촌경제연구원_품목관측 - 과일API
- <기상청>
 - 기상 자료 개방 포털

2. 데이터 수집 및 전처리

2. 데이터 전처리

The diagram illustrates the process of data preprocessing. It shows three separate data tables: A, B, and P. Table A has columns labeled A and B, with rows 1 through 3 containing data. Table B has columns labeled A and B, with rows 7 through 9 containing data. Table P has columns labeled P and 수박 (Watermelon), with rows 1 through 3 containing data. Ellipses between the tables indicate they represent multiple data points or time steps. A double-headed arrow connects the bottom table (B) to the rightmost table (P), suggesting a relationship or transformation between them.

	A	B
1		감귤
2	1	1.6
3	2	1.7
⋮	⋮	⋮
7	6	0.8
8	7	0.7
9	8	0.7
⋮	⋮	⋮
11	11	1.2
12	12	1.6
⋮	⋮	⋮

P	수박
1	0.7
2	0.7
3	
⋮	⋮
	1.6
	1.7
	1.7
⋮	⋮
	0.9
	0.8

1. 과일 데이터로드 → 필요칼럼 선별 → 결측치 확인
2. 없는 파일들은 초기 baseline 설정
3. 과일 별 시즌 계수, 명절 패턴 반영해 추정값 삽입

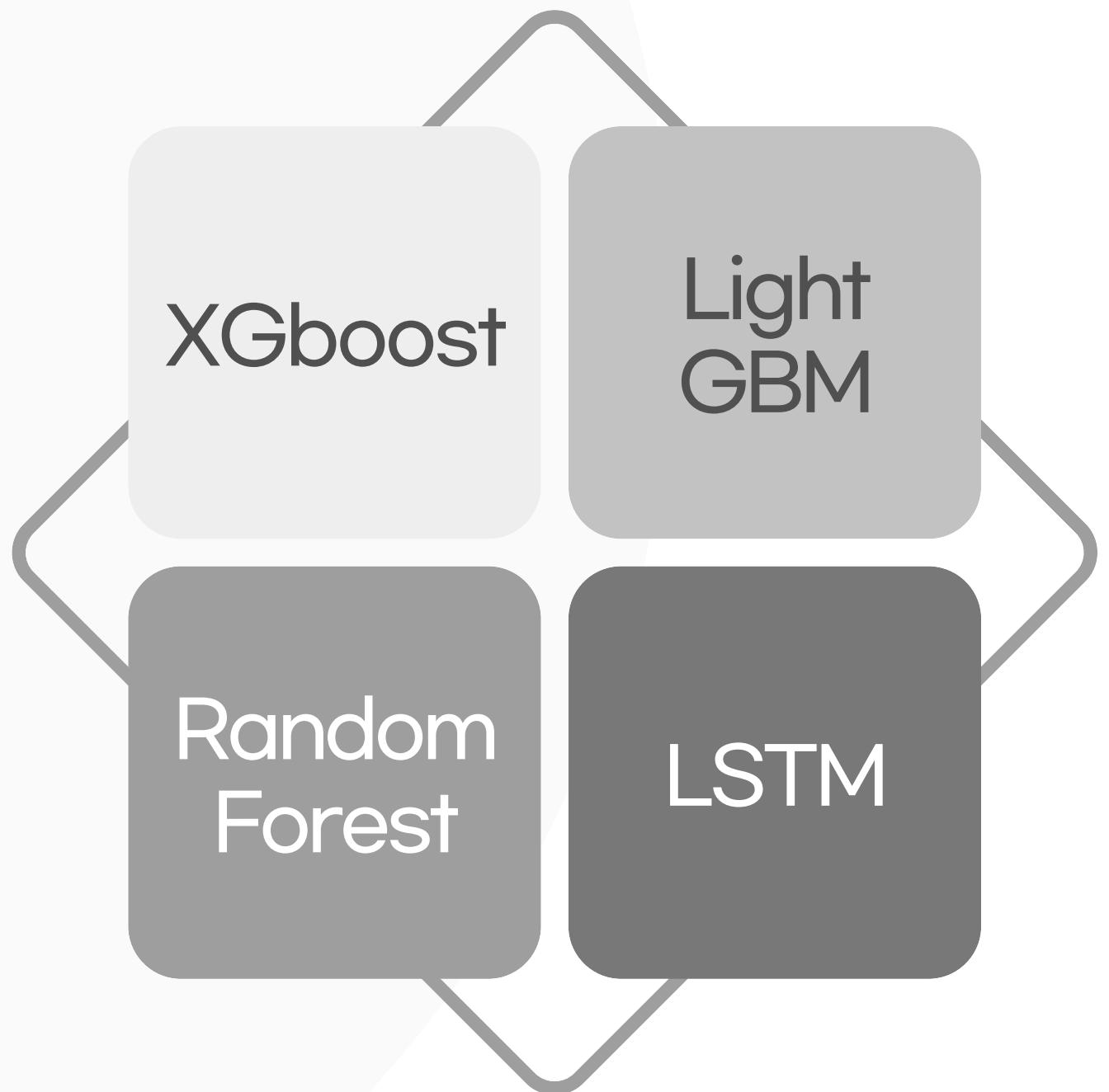


시즌 별 제철 과일

- 감귤류: 겨울 ↑, 여름 ↓
- 베리류: 2-4월 ↑
- 멜론류: 6-8월 ↑
- 열대과일 6-9월 ↑
- 사과 · 배: 8-10월 ↑
- 핵과류(복숭아, 자두, 살구 ...): 6-8월 ↑

3. 모델 설명 및 평가

0. 주요 모델 설명



XGBoost

목적: 예측 정확도를 높이기 위한 고성능 Gradient Boosting 모델 적용

특징: 트리를 순차적으로 학습하며 오차를 보완하여 높은 정확도와 규제가 포함돼 안정적

LightGBM

목적: XGBoost의 성능을 유지하면서 학습 속도 및 처리 효율 향상

특징: Leaf-wise 방식으로 미세한 패턴까지 포착하며 대규모 데이터도 빠르게 학습

Random Forest

목적: 머신러닝 기반 모델 중 안정적인 baseline 확보

특징: 여러 결정트리를 병렬 학습해 과적합을 줄이고 일관된 예측 성능 제공

LSTM

목적: 시계열 구조를 반영해 계절성, 추세 기반 미래 수요 변화까지 예측

특징: 과거 데이터를 기억해 순차적으로 학습하여 시간 의존성 반영 가능

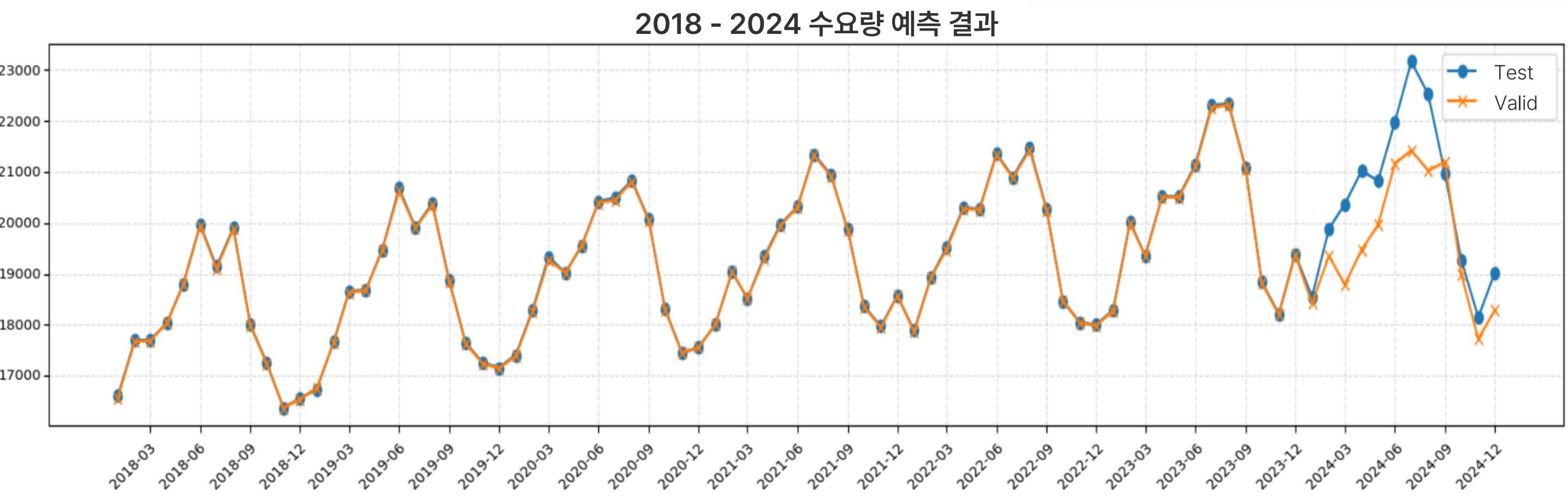
3. 모델 설명 및 평가

1. XGBoost

장점: 예측 정확도가 높고 복잡한 패턴 학습에 강함

단점: 데이터가 많을수록 학습 시간이 길고 튜닝이 복잡함

결론: 트리기반에서 안정적인 성능을 보여주며 정밀 예측에 활용 가능하지만 다른 모델에 비해 정확도는 약간 떨어지는 편



NO	Description
R ²	0.94
MAE	2878.70
RMSE	1987.29
학습시간	6m

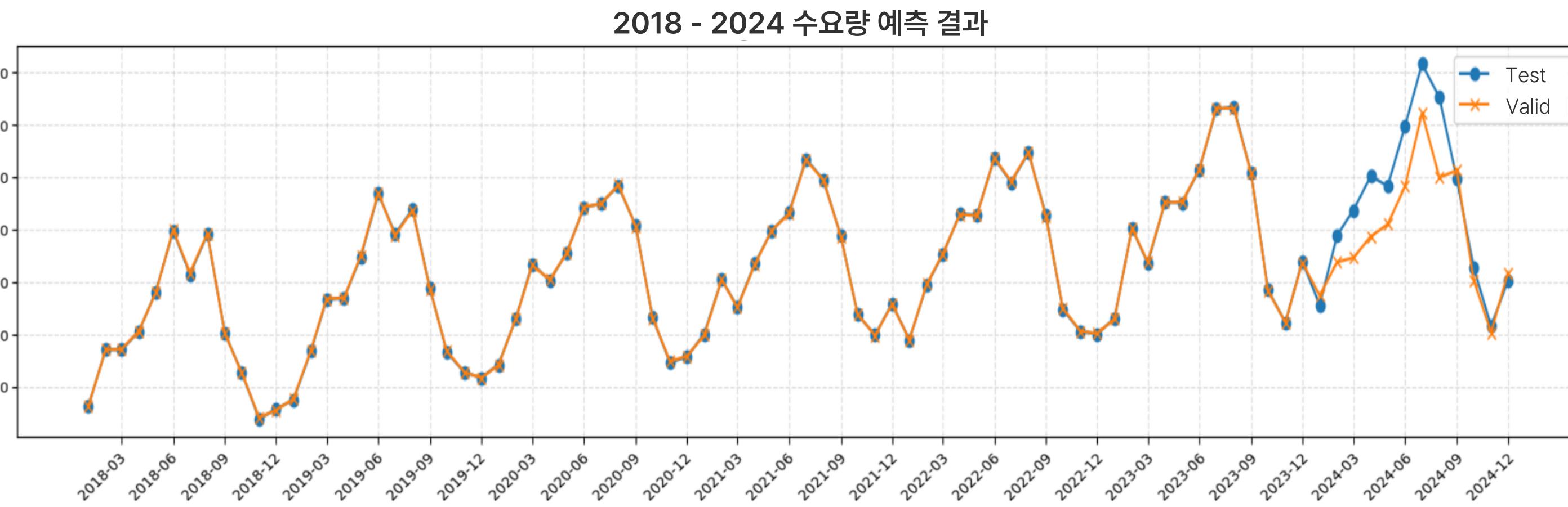
3. 모델 설명 및 평가

2. LightGBM

장점: 학습 속도가 가장 빠르고 효율적이며 정확도도 우수함

단점: 과적합 가능성이 존재함

결론: 학습 속도가 빠르고 정확도도 높아 가장 적합한 모델



NO	Description
R ²	0.95
MAE	1872.40
RMSE	2530.95
학습시간	1m 4s

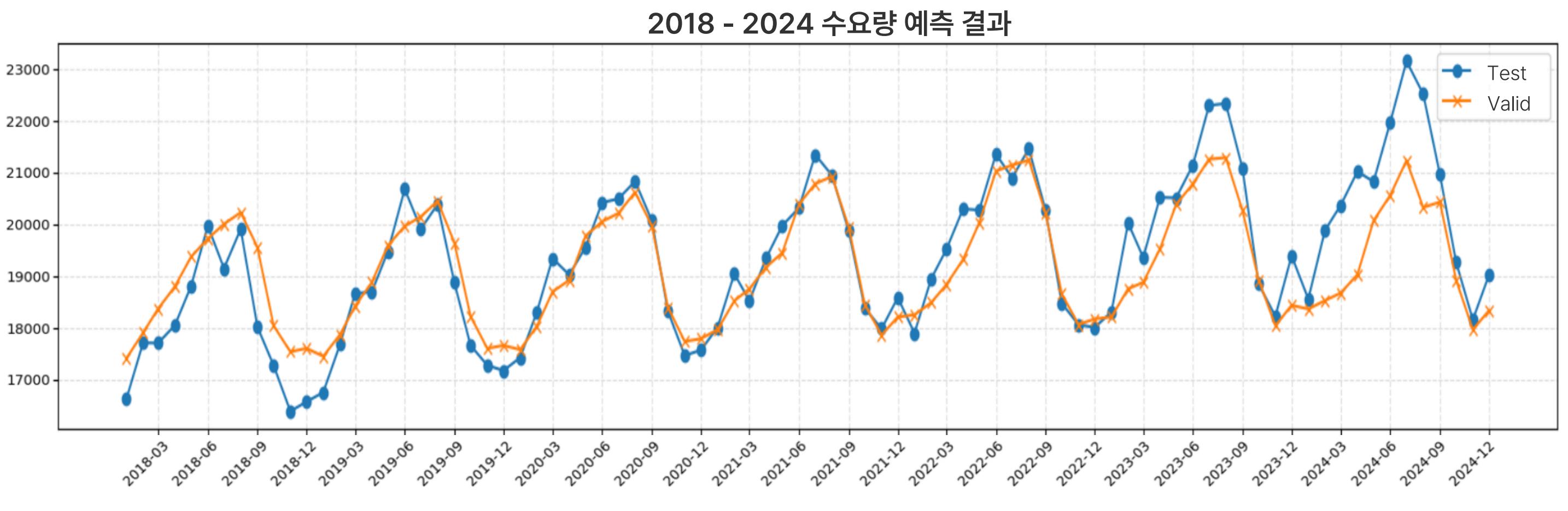
3. 모델 설명 및 평가

3. Random Forest

장점: 안정적인 예측과 해석이 용이함

단점: 세부 변화에 대응력이 낮고 학습이 비교적 느림

결론: 안정적이지만 세부 변동 포착이 부족해 베이스라인으로 적합함



NO	Description
R ²	0.92
MAE	2345.39
RMSE	3266.28
학습시간	7m

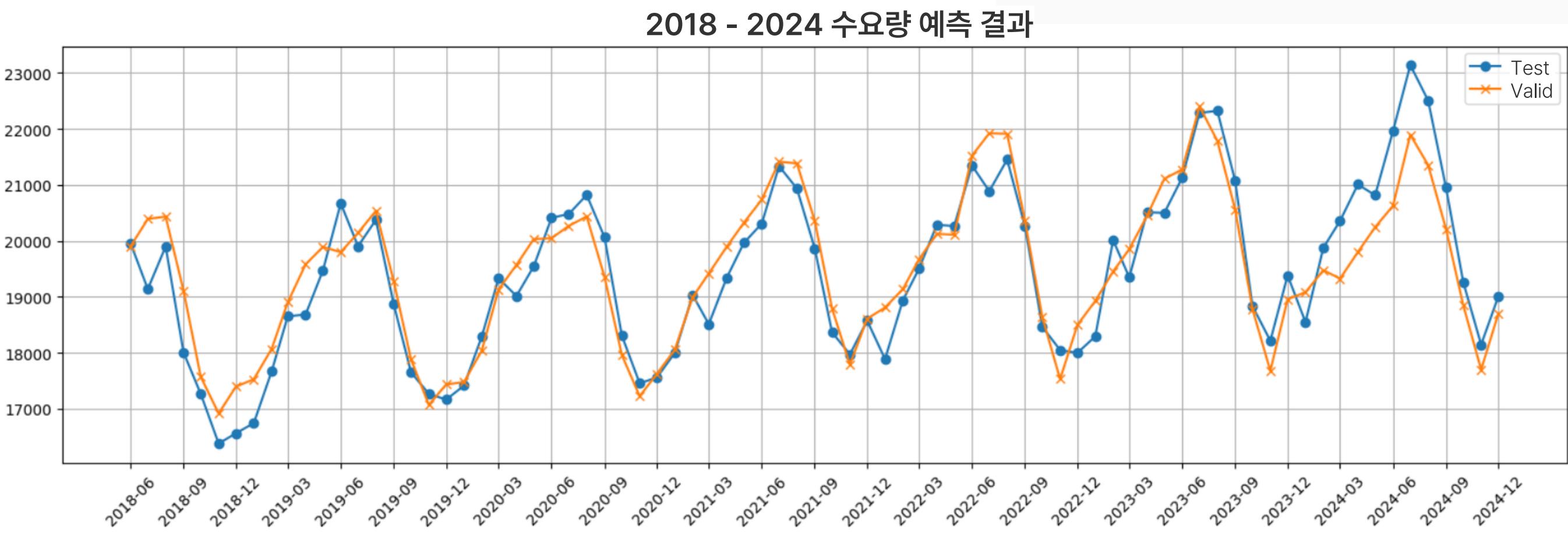
3. 모델 설명 및 평가

4. LSTM

장점: 시계열 패턴을 학습하는 능력이 좋고 실제 추세와 근접함

단점: 설계가 복잡하고 파라미터에 민감함

결론: 시계열 패턴을 가장 잘 반영하지만 몇몇 파일들은 정확도가 매우 낮음



NO	Description
R ²	0.96
MAE	1751.71
RMSE	2350.59
학습시간	1m 10s

3. 모델 설명 및 평가

5. 평가

4개의 모델을 비교한 결과, 전반적으로 정확도가 높고 편차가 크지 않은 성능을 확인 할 수 있음

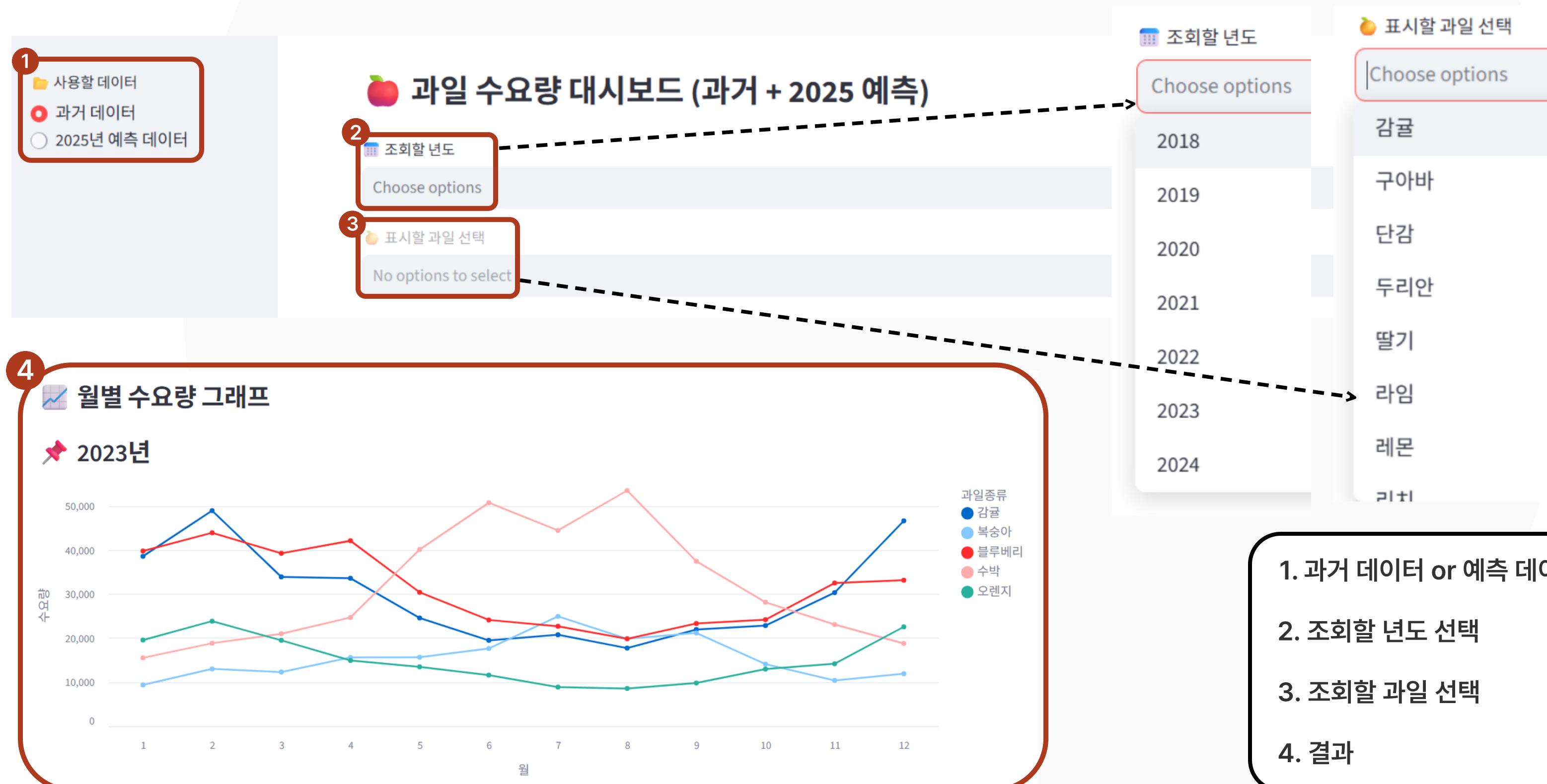
단감	-0.2496	553.7239	423.7937
발률토마토	-0.5880	1497.8454	1302.8706
석류	-0.9552	2119.6067	1499.0625
키위	-0.0168	1582.4508	1304.2187
토마토	-0.5761	1624.5960	1317.8032

속도 면에서는 LightGBM과 LSTM이 우수했으나 LSTM은 일부 파일에서 예측 정확도가 낮게 나타남

실제 사용자(파일가게 운영자, 도소매 업자 등) 데이터를 지속적으로 수정하거나 데이터 크기가 확장될 수 있다는 점을 고려했을 때 다양한 데이터 규모에 유연하게 대응할 수 있는 모델인 LightGBM을 선정

항목	XGBoost(6m)	LightGBM(1m04s)	Random Forest(7m)	LSTM(1m 10s)
R ²	0.94	0.95	0.92	0.96
MAE	2878.70	1872.40	2345.39	1751.71
RMSE	1987.29	2530.95	3266.28	2350.59

4. 웹 구현



4. 웹 구현

1. 사용할 데이터

- 사용할 데이터
- 과거 데이터
- 2025년 예측 데이터

2. 조회할 년도

Choose options

3. 표시할 파일 선택

No options to select

4. 상세 데이터

테이블 표시 과일

날짜	연도	월	과일종류	수요량
18	2025		1 사과	3090
57	2025		2 사과	3557
96	2025		3 사과	3340
135	2025		4 사과	4751
174	2025		5 사과	4622
213	2025		6 사과	5323
252	2025		7 사과	6844
291	2025		8 사과	5375
330	2025		9 사과	6865
369	2025		10 사과	5801

조회할 년도

Choose options

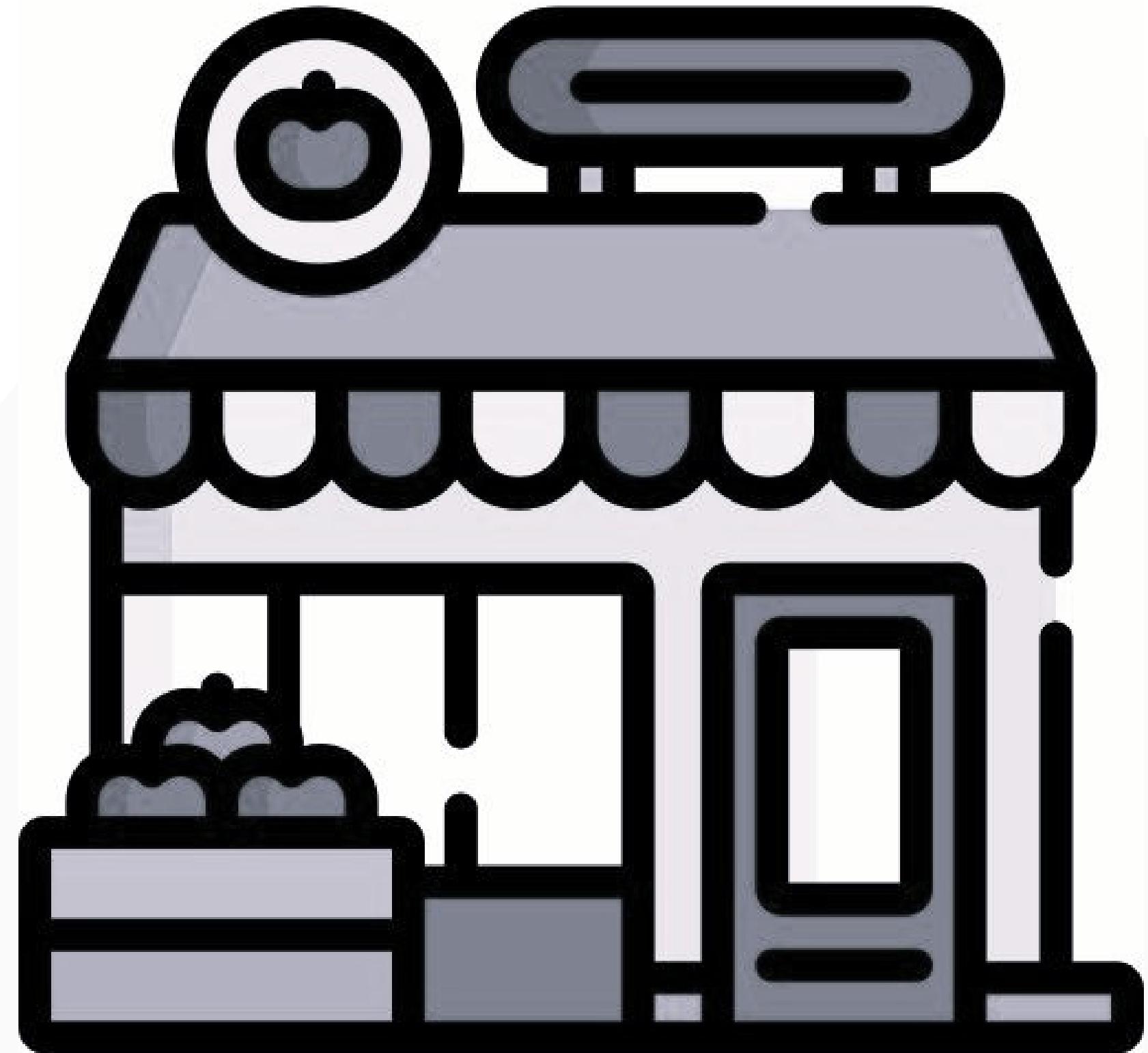
2018
2019
2020
2021
2022
2023
2024

1. 과거 데이터 or 예측 데이터 선택
2. 조회할 년도 선택
3. 조회할 파일 선택
4. 결과

5. 결론 및 평가

● 기대 효과

- 1. 재고 리스크 감소
과잉 재고 방지 → 폐기 비용 절감
재고 부족 예방 → 판매 손실 감소
= 수요·공급 균형 확보 & 운영 안정성 향상
- 2. 비용 효율 최적화
재고 보관·폐기 관련 비용 감소
적정 주문량 확보로 원가 효율 상승
- 3. 매출 증대 가능성
성수기·명절 수요 대응 → 판매량 극대화
수요 예측 기반 마케팅/랜딩 전략 활용 가능
- 4. 데이터 기반 의사결정 체계 구축
발주 판단을 감에서 데이터 기반으로 전환
AI 모델 고도화 → 장기적 자동 발주 시스템으로 확장



5. 결론 및 평가

성득명 (데이터 수집, Streamlit, 발표)

프로젝트를 진행하면서 모델을 구성하는 단계에서는 동원훈련으로 충분히 참여하지 못해 아쉬움이 컸습니다. 하지만 이후 스트림릿으로 대시보드를 만들고 허깅페이스에 배포하는 과정에서 많은 것을 배울 수 있었습니다. 스트림릿은 처음 사용해보았지만, 간단하게 프론트 화면을 구성할 수 있어 새로운 경험이 되었고 프로젝트 완성도에도 도움이 되었습니다.

김태훈 (모델 구성)

이번 AI 프로젝트는 처음엔 막막했지만, 직접 데이터를 다루고 모델을 만들어 보면서 AI가 어떻게 작동하는지 실감할 수 있었습니다. 특히 결과가 잘 나오지 않을 때의 답답함과 해결했을 때의 성취감이 강하게 남았습니다. 어려움도 있었지만 그만큼 많이 배우고 성장한 의미 있는 시간이었습니다.

이호 (데이터 수집, 모델 구성)

처음 프로젝트를 시작했을 때는 AI 부분이 너무 어려워 잘 해낼 수 있을까? 걱정도 됐지만, 직접 만들어 보고 부딪히면서 배우는 재미도 느낄 수 있었습니다. 진행하면서 부족한 부분을 다시 공부하고, 오류를 고치며 이해가 깊어지는 경험을 했습니다. 쉽지 않았지만 팀원들과 함께 해결해 가며 성장할 수 있었고, 그 과정 자체가 큰 배움이었습니다. 앞으로도 배운 내용을 직접 적용하며 더 발전하고 싶습니다. 함께한 팀원들에게 고맙다는 말을 전하고 싶습니다.

박선영 (데이터 수집, 모델 구성, PPT)

이번 프로젝트를 통해 수업 시간에 배웠던 내용을 직접 적용하며 이해가 부족했던 부분을 알아가고 자연스럽게 복습할 수 있었습니다. 또한 다양한 모델을 직접 구현해보며 이론으로만 배웠던 내용이 실제 데이터에서 어떻게 작동하는지 확인하고 경험할 수 있었습니다. 이를 통해 확실히 이해가 더 되었고 다음에 비슷한 프로젝트를 진행한다면 지금보다 더 수월하게 모델을 구축할 수 있을 것 같습니다.