

# AI 기반 판매 수요 예측

아시아경제 빅데이터 기반 비즈니스 모델 맞춤형 AI 솔루션 개발

Pulasia

푸라시아



# 목차

## 01 프로젝트 개요

- 1.1 팀 소개
- 1.2 프로젝트 목표
- 1.3 간트 차트

## 02 시스템 설계

- 2.1 시스템 아키텍처
- 2.2 개발 환경

## 03 데이터 분석 및 모델링

- 3.1 데이터 개요
- 3.2 데이터 전처리
- 3.3 모델링 프로세스

## 04 웹페이지 구현 및 시연

- 4.1 화면 설계
- 4.2 메뉴 구조도
- 4.3 UI 화면
- 4.4 데모 시연





# 팀 소개



팀장  
이인기



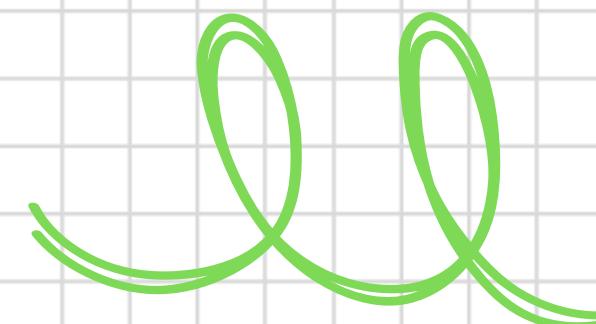
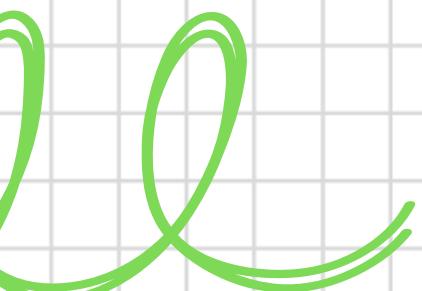
팀원  
전유진



팀원  
이경민



멘토  
김성희



# 프로젝트 목표

## 데이터 분석

- 판매 데이터를 기반으로 과거 판매량 분석 시각화

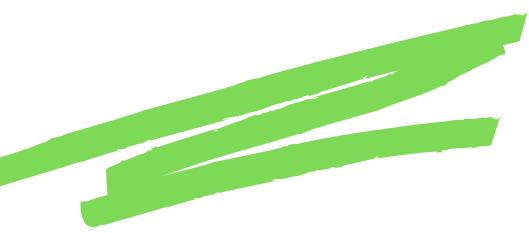
## 수요 예측

- 실제 데이터를 기반으로 수요 예측
- 다양한 딥러닝 시계열 분석 모델 사용

## 웹 개발 및 데이터 통신

- 분석한 데이터를 웹을 통해 표현
- API를 통한 서로 다른 포트 간 데이터 통신

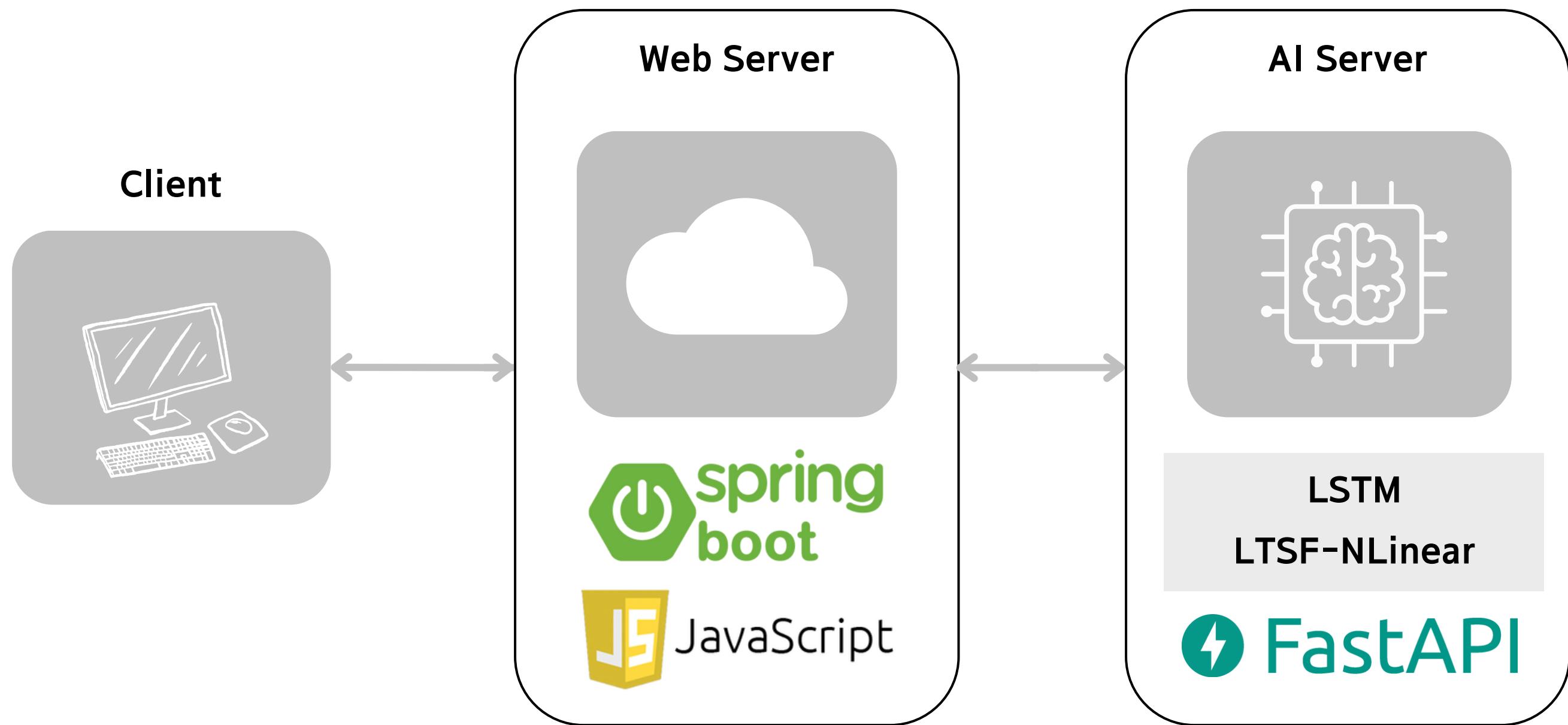




# 간트차트



# 시스템 아키텍처



# 개발 환경



# 데이터 개요

## ■ 원천 데이터

Rows 10,000,000

Columns 6

## ■ 컬럼 설명

Unnamed: 0 고객ID

date 날짜/시간

cate 상품군

name 상품명

mart 지점

tot 판매 수량

	Unnamed: 0	date	cate	name	mart	tot
0	307604	2020-01-01 13:32:17	Eggplant Oriental		NaN	2 7
1	532169	2020-01-01 13:32:37	Ham	Cooked Italian	8	4
2	887328	2020-01-01 13:36:51	Cheese	Parmesan Cubes	9	3
3	66843	2020-01-01 13:36:55	Bay Leaf		NaN	2 4
4	929967	2020-01-01 13:37:26	Bread Fig And Almond		NaN	8 11

[원천 데이터 상위 5개 행]

# 데이터 전처리



결측치 처리

ADF 검정

데이터 스케일링

데이터로더 구성



맞춤법/띄어쓰기 교정

pyspellchecker를 이용한  
명사 표준화

명사 단·복수 일치화

정규표현식을 이용한  
단수화

공백 제거

LTRIM을 이용한  
상품명 왼쪽 공백 제거

상품군·상품명 표준화

# 데이터 전처리



상품명 결측치를 상품군 값으로 대체

	date	mart	tot	cate	name		date	mart	tot	cate	name
0	2020-01-01	2	7	Eggplant Oriental	NaN	0	2020-01-01	2	7	Eggplant Oriental	Eggplant Oriental
1	2020-01-01	8	4	Ham	Cooked Italian	1	2020-01-01	8	4	Ham	Cooked Italian
2	2020-01-01	9	3	Cheese	Parmesan Cubes	2	2020-01-01	9	3	Cheese	Parmesan Cubes
3	2020-01-01	2	4	Bay Leaf	NaN	3	2020-01-01	2	4	Bay Leaf	Bay Leaf
4	2020-01-01	8	11	Bread	Fig And Almond	4	2020-01-01	8	11	Bread	Fig And Almond

# 데이터 전처리

텍스트 정제

결측치 처리

ADF 검정

데이터 스케일링

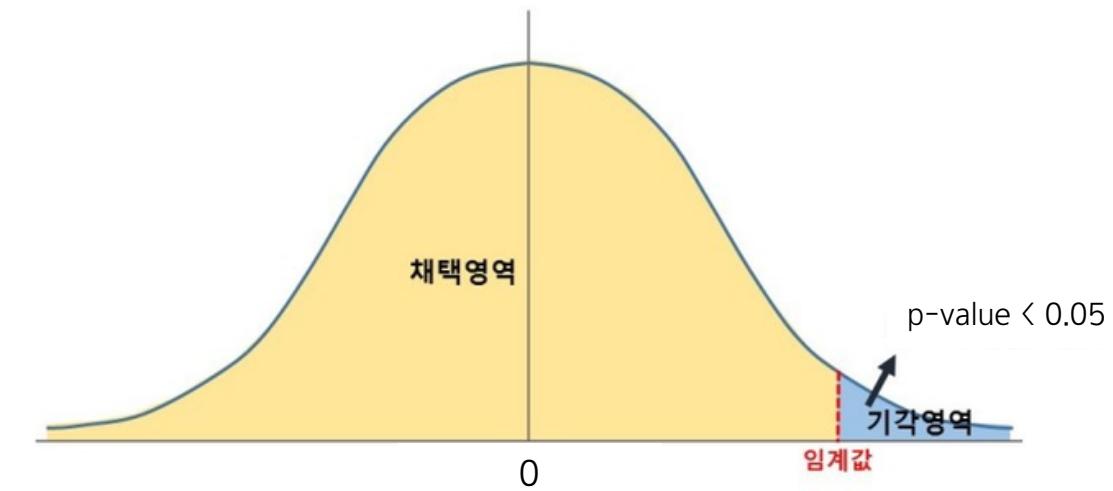
데이터로더 구성



## ADF 검정

단위근 존재 여부를 통해 시계열 데이터가 다음 조건을 갖춘 **정상성** 데이터인지 확인

“정상성이란,  
데이터가 시간과 관계없이 일정한 성질을 가지는 것”



ADF Statistic: -2.6439776603203966

p-value: 0.08424427568585385

- ▶ p-value  $\geq 0.05$  귀무가설 채택 → **비정상 시계열**
- ▶ 시계열 데이터에 **추세나 계절성**과 같은 변화 존재

# 데이터 전처리

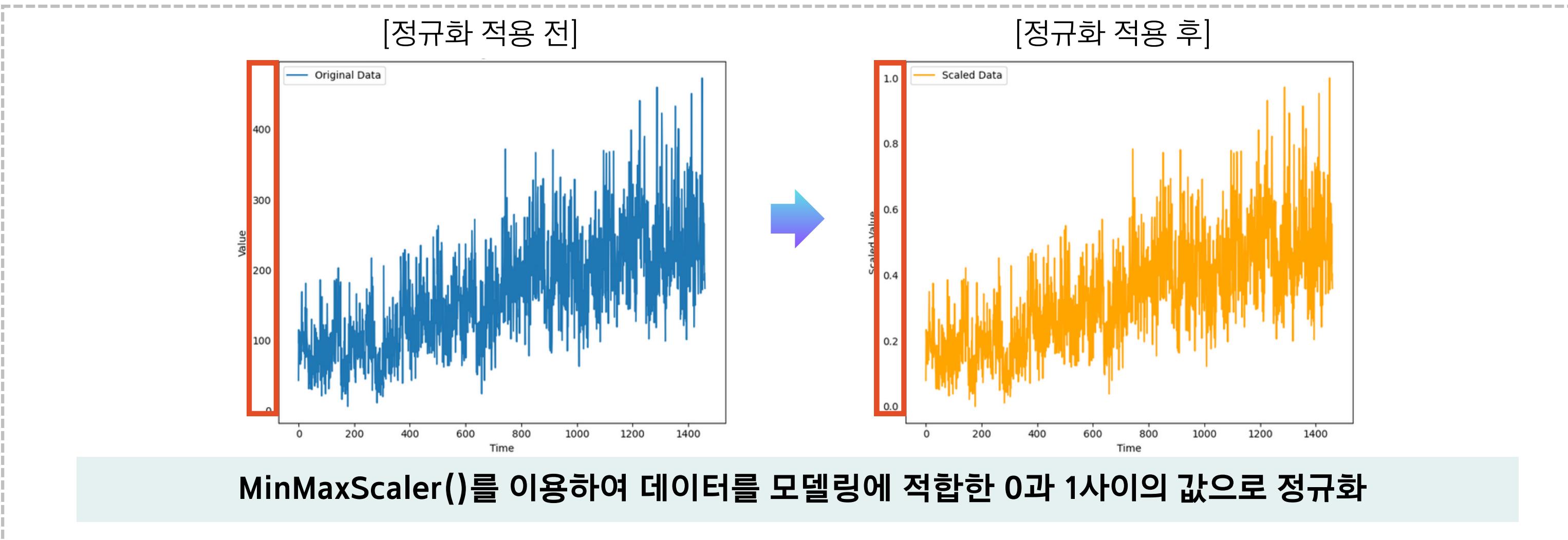
텍스트 정제

결측치 처리

ADF 검정

데이터 스케일링

데이터로더 구성



# 데이터 전처리



✓ PyTorch의 DataLoader를 사용하여 입력 데이터 로드

✓ 파라미터 입력

(scaled\_data, window\_size, forecast\_size, batch\_size)

✓ windowDataset 클래스 인스턴스 생성

모델이 학습 가능한 형태로 데이터 변환

# 모델링 프로세스

## 모델 설정

## 모델 학습

## 모델 평가 및 예측

## 결과물 시연

머신러닝

Light GBM (Tree)

딥러닝

LSTM (RNN)

LTSF-Linear

NLinear

DLinear

피처 불충분,  
시계열 특성 미고려

장기 의존성 우수,  
예측 모델로 선정

비선형, 추세 모델링에 적합  
예측 모델로 선정

비선형 패턴에 부적합

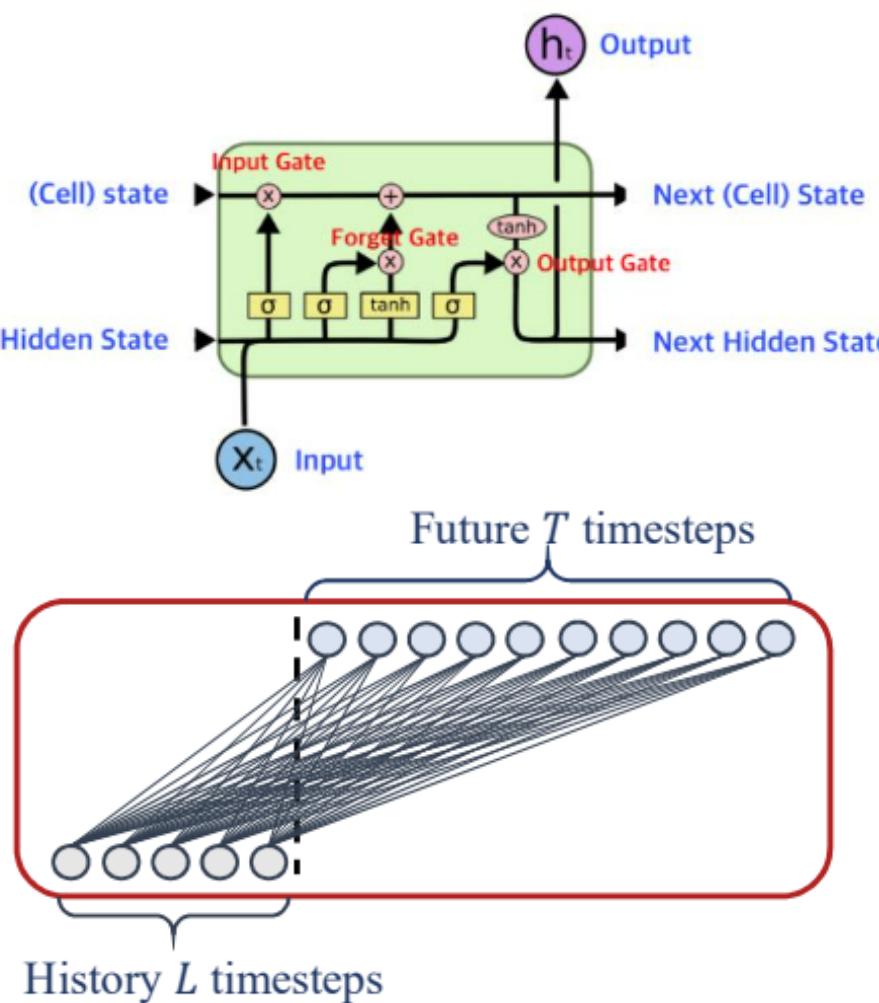
# 모델링 프로세스

## 모델 설정

## 모델 학습

## 모델 평가 및 예측

## 결과물 시연



## LSTM

- 우수한 장기 의존성 모델링 > 장기 시계열 데이터 패턴 파악에 적합
- 장점: 시퀀스 데이터의 복잡한 패턴 모델링, 가변적 입력 데이터 처리 가능
- 단점: 모델 학습 및 추론 속도 부진, 데이터가 긴 경우 많은 메모리 소비

## LTSF-Linear

- LSTM과 선형회귀의 장점 결합, 장단기 시계열 예측 성능 향상
- 장점: 우수한 예측 성능, 구현과 튜닝이 쉬운 간단한 모델 구조
- 단점: LSTM에 비해 장기 패턴 모델링 저성능

# 모델링 프로세스

모델 설정

모델 학습

모델 평가 및 예측

결과물 시연



Train 360days

Forecast 10days

Train 360days

Forecast 10days

Train 360days

Forecast 10days

360일 동안의 데이터 학습, 이후 10일 데이터 예측

# 모델링 프로세스

모델 설정

모델 학습

모델 평가 및 예측

결과물 시연



[Bagel 상품군의 모델별 시계열 예측 성능]

	MAE	RMSE	MAPE
LTSF-NLinear	35.78	52.6	18.68
LSTM	43.13	46.26	20.71

[예측 모델별 성능 비교]

MAE : LTSF-NLinear > LSTM  
RMSE: LSTM > LTSF-NLinear  
MAPE: LTSF-NLinear > LSTM

종합적으로, Bagel 상품군의 경우  
LTSF-NLinear 모델의 성능이 더 우수

실제 데이터와 예측 데이터를 비교한 예측 성능 평가

# 모델링 프로세스

모델 설정

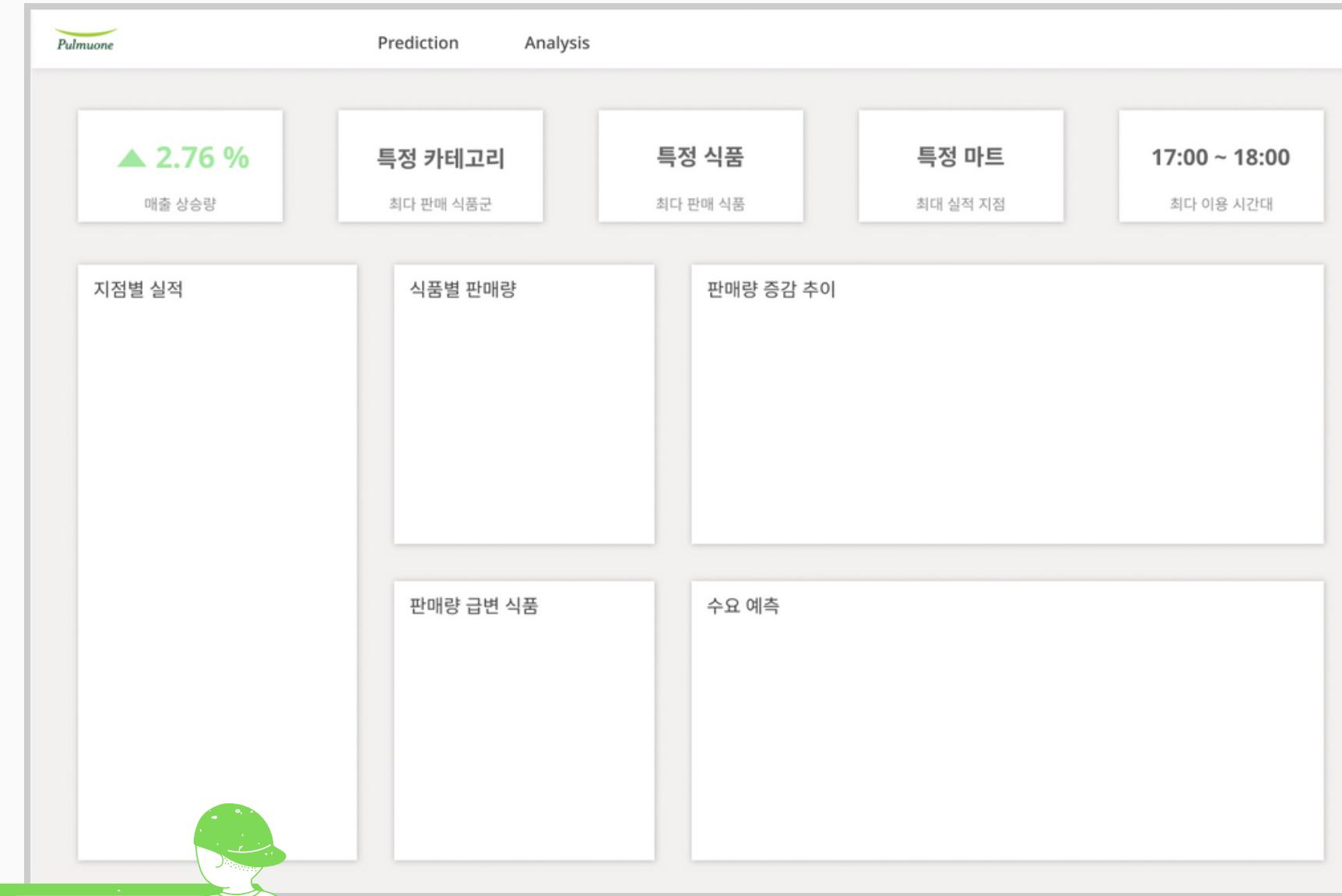
모델 학습

모델 평가 및 예측

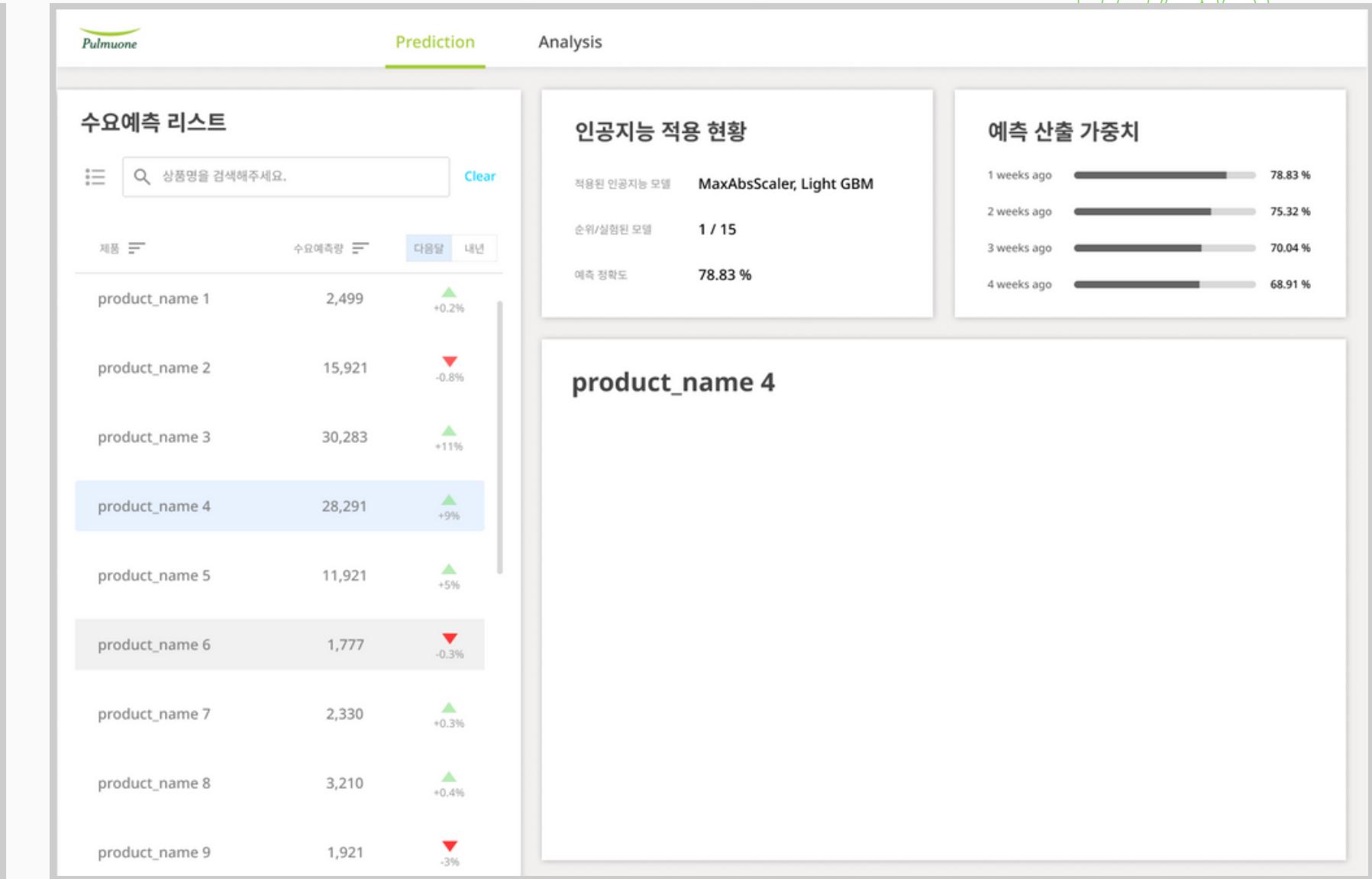
결과물 시연



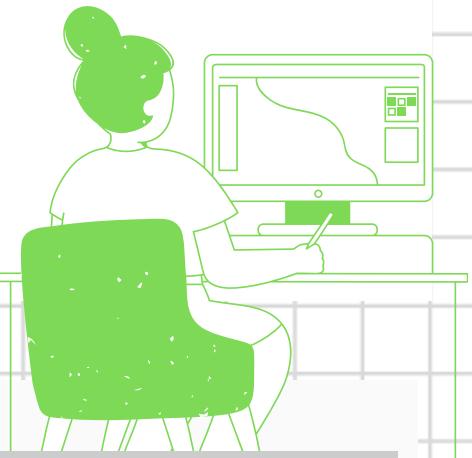
# 화면 설계



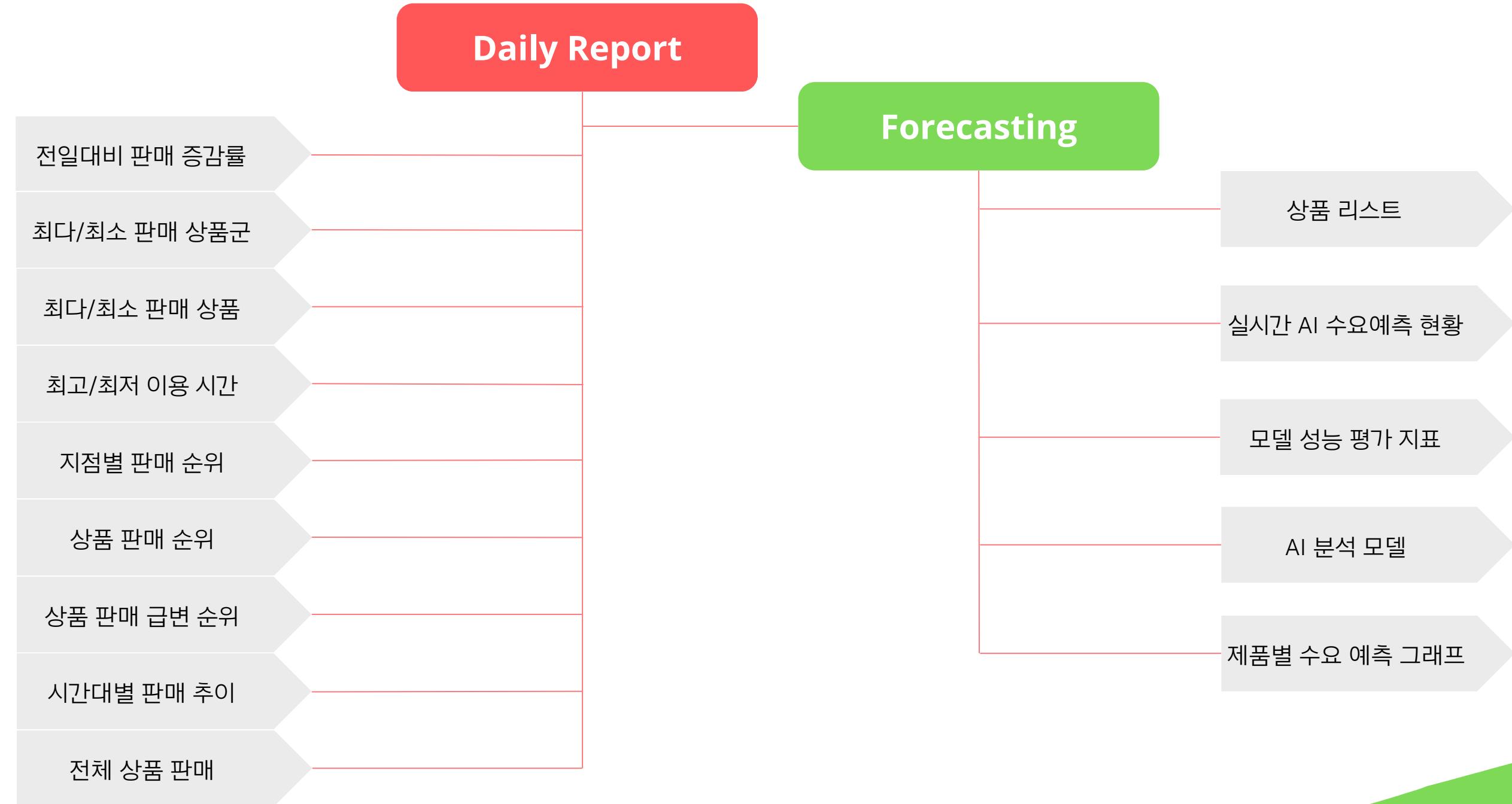
[분석 페이지 화면]



[수요 예측 페이지 화면]



# 메뉴 구조도



# UI 화면

## ■ 분석 화면

Pulmuone Daily Report Forecasting 2023년 12월 31일

**▲ 0.41%**  
전일대비 판매 증감률

**Wine**  
최다 판매 상품군

**Relish**  
최소 판매 상품군

**Whole Wheat**  
최다 판매 상품

**Parmesan Cubes**  
최소 판매 상품

**12:00 ~ 13:00**  
최고 이용 시간

**23:00 ~ 00:00**  
최저 이용 시간

**지점별 판매 순위**

지점	판매량	증감률
4	8544	(12.16%)
8	8271	(11.77%)
5	8081	(11.50%)
9	8064	(11.48%)
7	7680	(10.93%)
2	7525	(10.71%)
3	7520	(10.71%)
6	7362	(10.48%)
1	7197	(10.25%)

**상품 판매 순위**

순위	상품군	상품	판매량
1	Flour	Whole Wheat	423
2	Beets	Mini Golden	420
3	Pancetta	Pancetta	409
4	Papaya	Papaya	381
5	Duck	Legs	366

**시간대별 판매 추이**

**상품 판매 급변 순위**

순위	상품군	상품	전일	금일	판매 증감률
1	Soup	Tomato Mush. Florentine	43	284	▲ 560.47%
2	Chocolate	Pistoles, Lactee, Milk	33	159	▲ 381.82%
3	Dragon Fruit	Dragon Fruit	35	161	▲ 360.00%
4	Muffin	Banana Nut Individual	46	197	▲ 328.26%
5	Cake	Sheet Strawberry	45	182	▲ 304.44%

**전체 상품 판매**

# UI 화면

## ■ 수요 예측 화면

### (1) 초기 화면

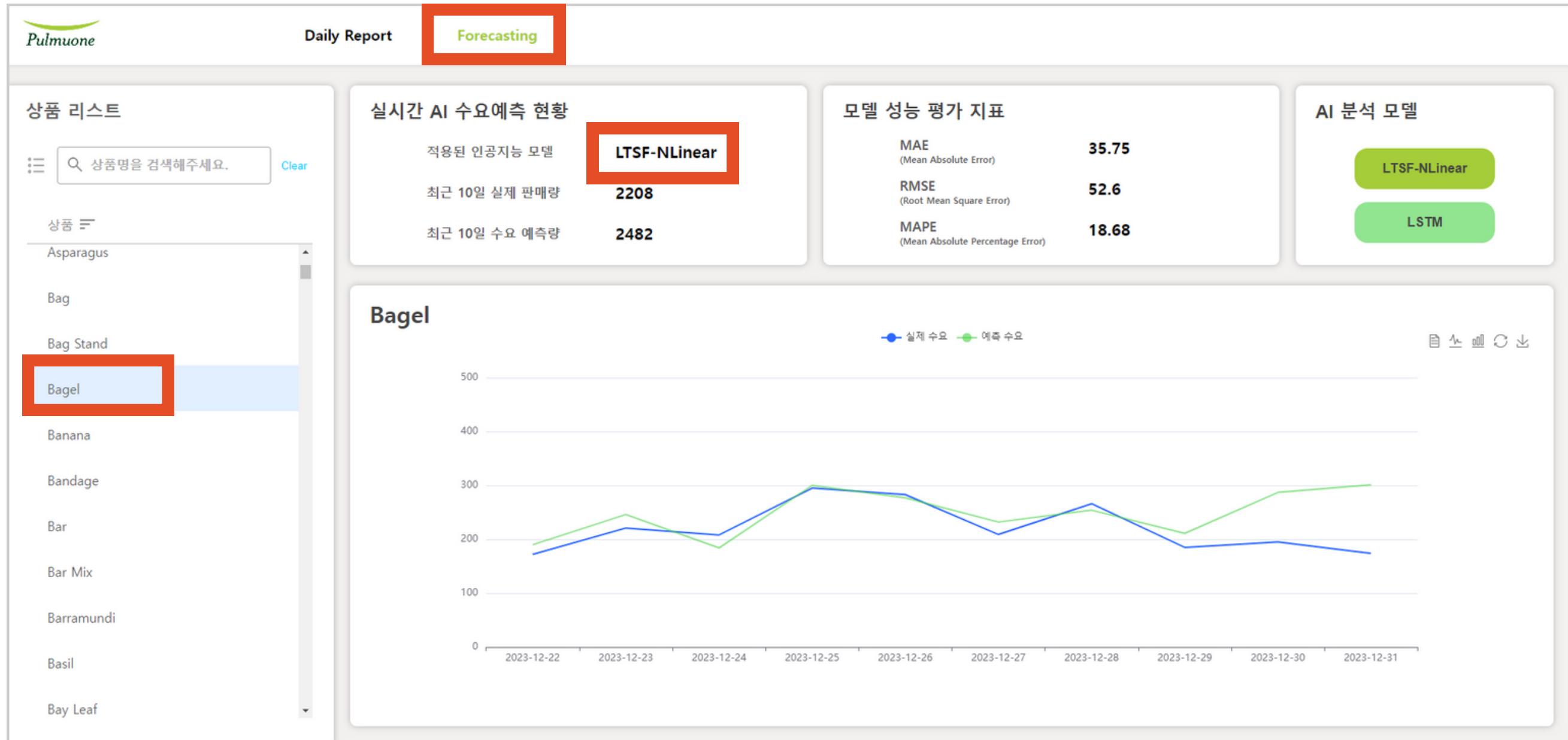
The screenshot shows a web-based forecasting application for Pulmuone. At the top, there is a navigation bar with three tabs: 'Pulmuone' (logo), 'Daily Report', and 'Forecasting'. The 'Forecasting' tab is highlighted with a red box. Below the navigation bar, there are four main sections:

- 상품 리스트**: A search bar with placeholder text '상품명을 검색해주세요.' and a 'Clear' button. Below it is a small input field labeled '상품'.
- 실시간 AI 수요예측 현황**: Displays three metrics:
  - 적용된 인공지능 모델 (Analysis Model)
  - 최근 10일 실제 판매량 (Recent 10-day Actual Sales)
  - 최근 10일 수요 예측량 (Recent 10-day Demand Forecast)
- 모델 성능 평가 지표**: Shows performance metrics with numerical values:
  - MAE (Mean Absolute Error)
  - RMSE (Root Mean Square Error)
  - MAPE (Mean Absolute Percentage Error)
- AI 분석 모델**: Buttons for selecting analysis models:
  - LTSF-NLinear
  - LSTM

At the bottom left, there is a large input field labeled '상품명' (Product Name) for searching specific products.

# UI 화면

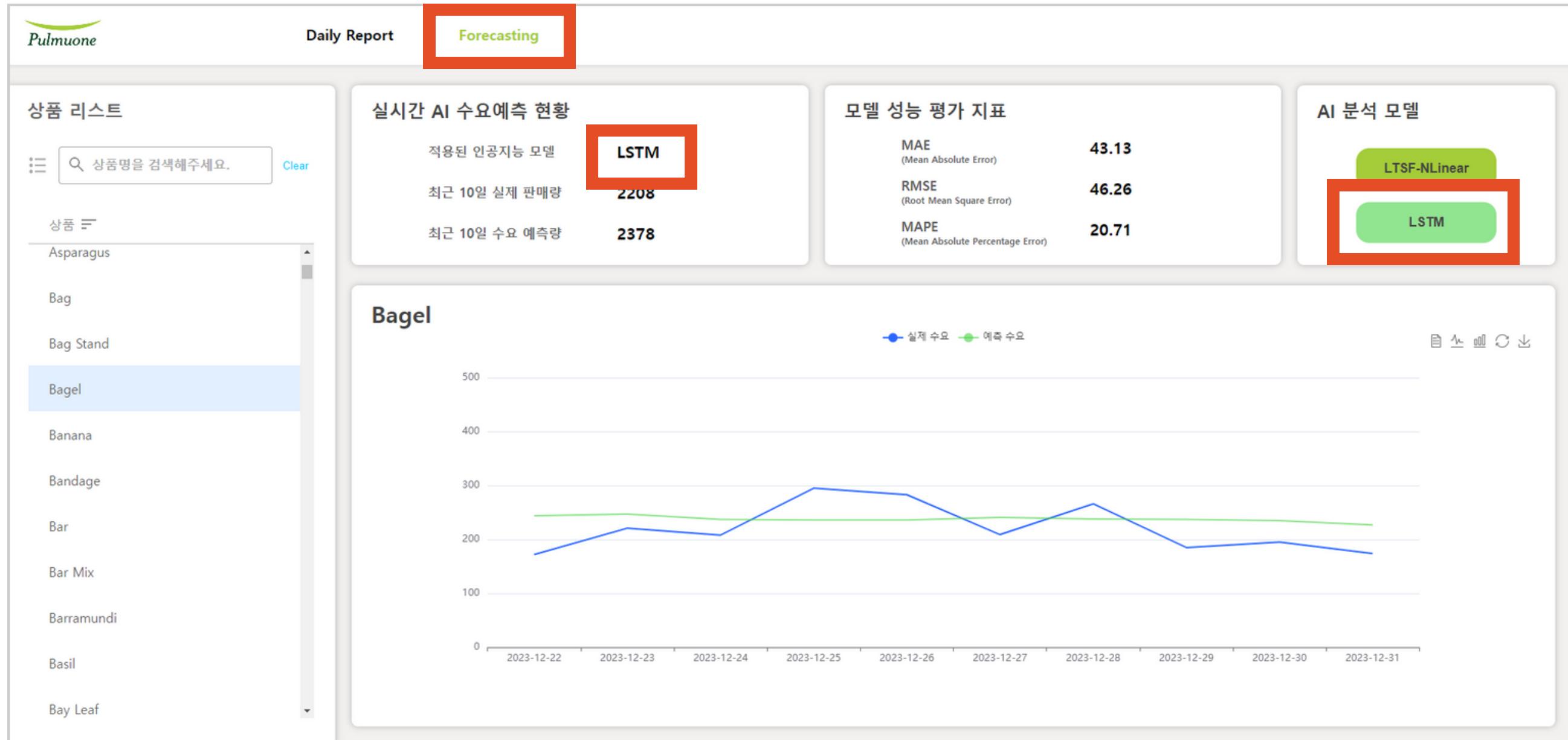
## ■ 수요 예측 화면 (2) 상품 선택



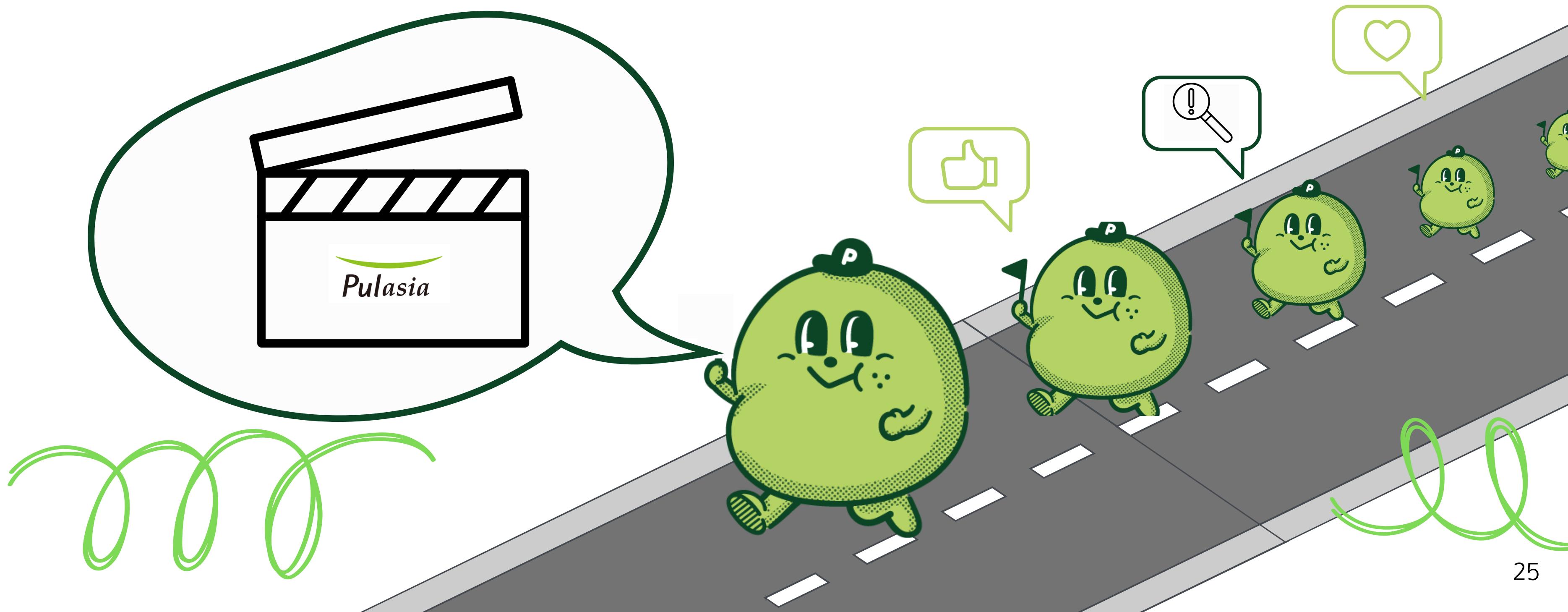
# UI 화면

## ■ 수요 예측 화면

### (3) LSTM 모델 선택 (모델 변경)



# 데모 시연



# Q & A

