

---

# 2023 POSTECH OIBC CHALLENGE

---

仁義禮智信



## **1. INTRO**

- 1-1. Background
- 1-2. Problem Definition

## **2. EDA**

- 2-1. Feature Analysis
- 2-2. Incentive Analysis

## **3. DESIGN**

- 3-1. System Architecture
- 3-2. Loss Function
- 3-3. Performance Metrics

## **4. IMPLEMENT**

- 4-1. Model Tuning
- 4-2. Model Selection

## **4. OUTRO**

- 5-1. Results
- 5-2. Conclusion

# INTRO

01

# Background. 좋은 양상불이란?

## 상황별 양상불



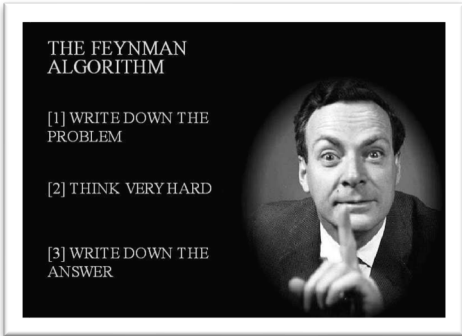
상황에 따른 유동성 고려  
→ 상황별 가중치 변화

## 데이터 기반 의사결정



확장성 고려  
→ 객관적인 척도 기반

## 시스템 최적화



해당 시스템 이해  
→ 목표 명세화 및 달성

상황별 앙상블

# “최적의 앙상블 기법”

시스템 최적화

WHAT?

“시스템 최적화” 관점에서 “집단 지성” 활용

HOW?

상황별 앙상블 특성 고려

확장성 고려

해당 시스템 이해

→ 상황별 가중치 배분 → 각 분야 전문가들의 협력과 경쟁 → 손실 최소화 관점의 의사결정 달성

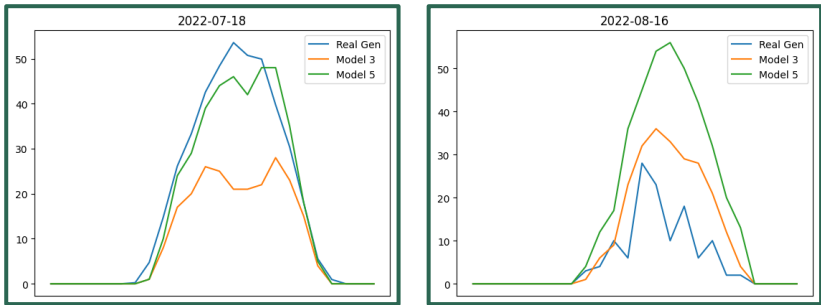
# EDA

# 02

A dark green horizontal bar at the bottom of the slide contains several light green geometric shapes: a circle on the left, followed by three squares of increasing size, and two rectangles of increasing size, all arranged in a diagonal line from bottom-left to top-right.

# Feature Analysis. 날씨 예측값 & 모델 예측값

날씨 예측값 반영 배경



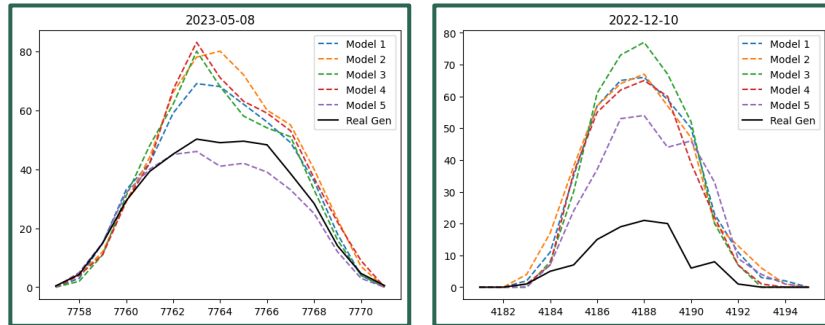
모델 5가 모델 3보다  
실제 발전량에 가깝게 예측

모델 3이 모델 5보다  
실제 발전량에 가깝게 예측

날씨(상황)별 개별 모델 성능이 상이함

날씨에 적응하여 모델에 매번 다른 가중치 할당

모델 예측값 반영 배경



실제 발전량이  
각 모델 발전량의 내분점에 해당

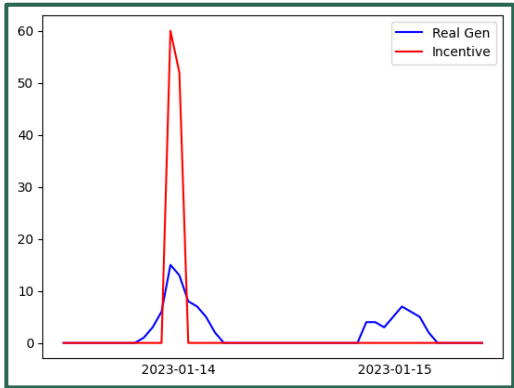
실제 발전량이  
각 모델 발전량의 외분점에 해당

실제 발전량을 각 모델 발전량의  
선형 결합으로 도출할 수 있음

각 모델 예측의 선형 결합으로 최적점 도달 가능

# Incentive Analysis. 실제 발전량 & 인센티브량

정량적 손실 반영 배경



일자/시간별 확보 가능한 최대 인센티브가 상이함

획득 가능 최대 인센티브 고려 필요

정성적 손실 반영 배경



평균오차와 확보 가능한 최대 인센티브가 상이함

임계오차율을 상회하는 예측에 집중 필요



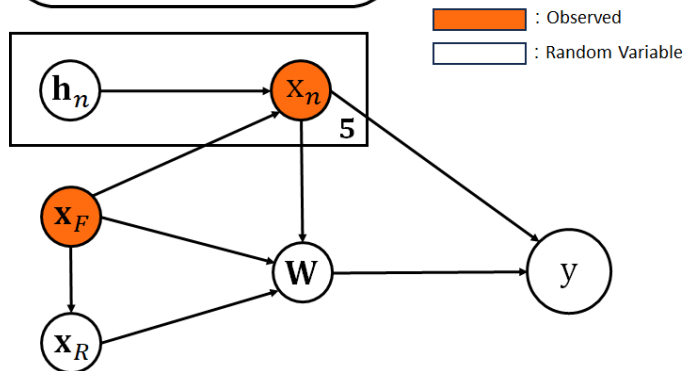
# DESIGN

03

# System Architecture. 실제 시스템 분석 & 아키텍처 설계

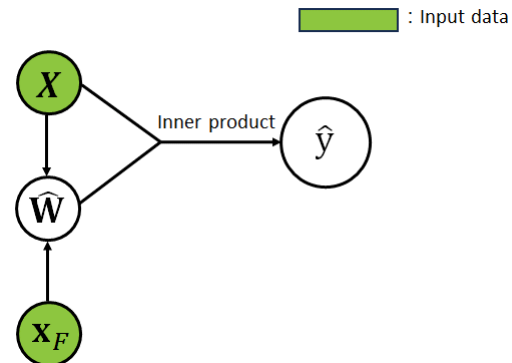
실제 시스템 그래프 모델

$h_n$ : 모델 n의 입력 변수  
 $x_n$ : 모델 n의 발전량 예측값  
 $x_F$ : 날씨 예측값  
 $x_R$ : 날씨 실측 변수  
 $W = \{w_n\}_{n=1}^5$   
 $w_n$ : 모델 n의 가중치  
 $y$ : 실제 발전량,  $\sum_{n=1}^5 w_n \cdot x_n$



아키텍처 제안

$X = \{x_n\}_{n=1}^5$   
 $x_n$ : 모델 n의 발전량 예측값  
 $x_F$ : 날씨 예측값  
 $\hat{W} = \{\hat{w}_n\}_{n=1}^5$   
 $\hat{w}_n$ : 모델 n의 추정 가중치  
 $\hat{y}$ : 앙상블 결과값,  $\sum_{n=1}^5 \hat{w}_n \cdot x_n$



# Loss Function. 손실함수 설계

## 손실함수

for dataset  $D := (x_i, y_i)_{i=1}^n$

quantitative loss ( $L_{quan}$ )  $:= c_i y_i$

$$\text{where } c_i = \begin{cases} 4 & \text{if } |f(x_i) - y_i| \times \frac{100}{99} > 8 \text{ and } y_i \geq 9.9 \\ 1 & \text{if } 6 < |f(x_i) - y_i| \times \frac{100}{99} \leq 8 \text{ and } y_i \geq 9.9 \\ 0 & \text{otherwise} \end{cases}$$

qualitative loss ( $L_{qual}$ )  $:= |f(x_i) - y_i|^\alpha$

( $\alpha$  : hyperparameter,  $0 < \alpha < 1$ )

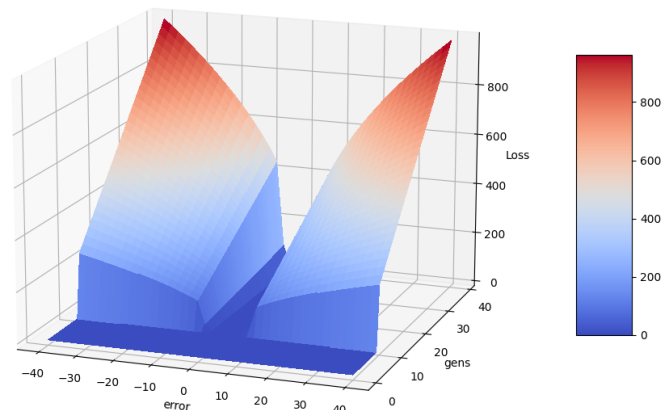
$$L = L_{quan} \cdot L_{qual}$$

$$\text{i.e. } L = \frac{1}{n} \sum_{i=1}^n c_i y_i |f(x_i) - y_i|^\alpha$$

$f(x_i)$  : 태양광 발전소  $i$  시간대 앙상블 결과값

$y_i$  : 태양광 발전소  $i$  시간대 실제 발전량

## 손실함수 시각화



성능지표

$$\text{efficiency} = \left( 1 - \frac{\sum_{i=1}^n c_i y_i}{\sum_{i=1}^n 4I_i y_i} \right) \times 100\%$$

where  $c_i = \begin{cases} 4 & \text{if } |f(x_i) - y_i| \times \frac{100}{99} > 8 \text{ and } y_i \geq 9.9 \\ 1 & \text{if } 6 < |f(x_i) - y_i| \times \frac{100}{99} \leq 8 \text{ and } y_i \geq 9.9 \\ 0 & \text{otherwise} \end{cases}$

$$I_i = \begin{cases} 1 & \text{if } y_i \geq 9.9 \\ 0 & \text{otherwise} \end{cases}$$

$f(x_i)$  : 태양광 발전소  $i$  시간대 앙상블 결과값  
 $y_i$  : 태양광 발전소  $i$  시간대 실제 발전량

성능지표 해석

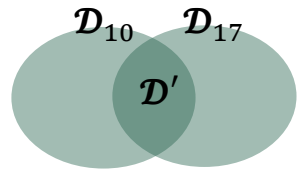
- 오차율이 전부 6% 이하인 경우  
100%
- 오차율이 전부 6% 초과 8% 이하인 경우  
75%
- 오차율이 전부 8% 초과인 경우  
0%

# IMPLEMENT

04

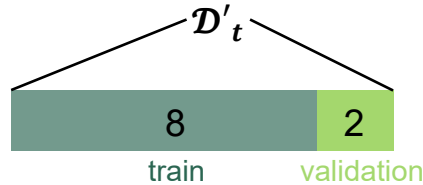
# Model Tuning. 모델 구조 결정

## 결측치 제거



10시와 17시 중 어느 하나라도 결측치가 존재하는 시간대 제거

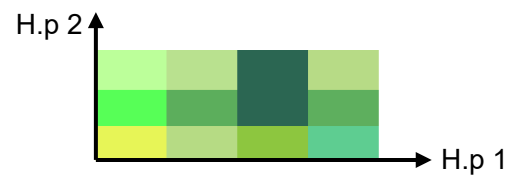
## 데이터 분할



$D'_t :=$  결측치가 제거된 데이터  $D'$ 의  $t$ 시 데이터 ( $t=10, 17$ )

데이터( $D'_t$ )를 8:2로 분할하여 각각 앙상블 모델  $t$ 의 학습과 검증에 사용

## 최적 세팅 탐색



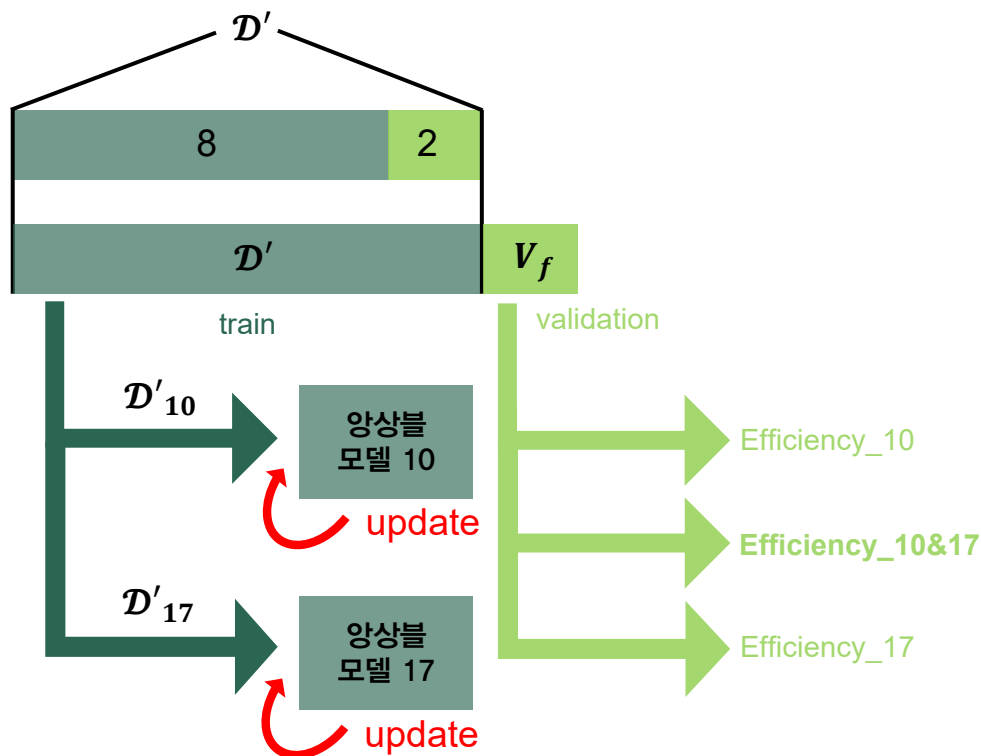
그리드 서치를 활용하여 앙상블 모델  $t(t=10,17)$ 의 최적 하이퍼파라미터 조합 탐색



<앙상블 모델  $t$  구조,  $t=10, 17$ >

- **Learning rate** =  $10^{-5}$
- **Batch size** = 128
- **Optimizer** = Adam
- $\alpha = \frac{1}{2}$
- **Structure of FCL**
  - model\_10 = (128, 128, 128, 128, 5)
  - model\_17 = (64, 64, 64, 64, 64, 5)

# Model Selection. 최적의 모델 조합 도출



## 검증을 통한 최적 모델 조합 도출

- 앞서 8:2로 train/validation split한 데이터를 합쳐서 앙상블 모델을 재학습
- 대회 사전참여 기간의 데이터(10/17-11/12)  $V_f$ 를 사용하여 학습과 동시에 모델을 검증
- 각 모델의 efficiency와 두 모델 예측의 평균을 예측값으로 할 때의 efficiency 계산
- 두 모델 예측 평균의 efficiency를 최대화하는 앙상블 모델 10과 앙상블 모델 17 선정

	앙상블 모델 10	앙상블 모델 17	앙상블 모델 10&17
Efficiency	57.59%	58.13%	61.00%

“평가 산식상 최대 인센티브의 61%를 획득하게끔 하는 앙상블 모델 10과 앙상블 모델 17 학습 완료 ”

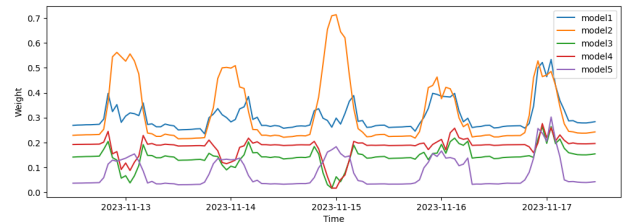
# OUTRO

05

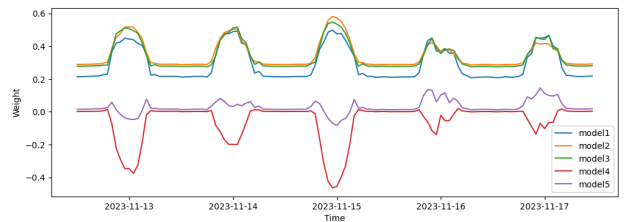


# Results. 결과

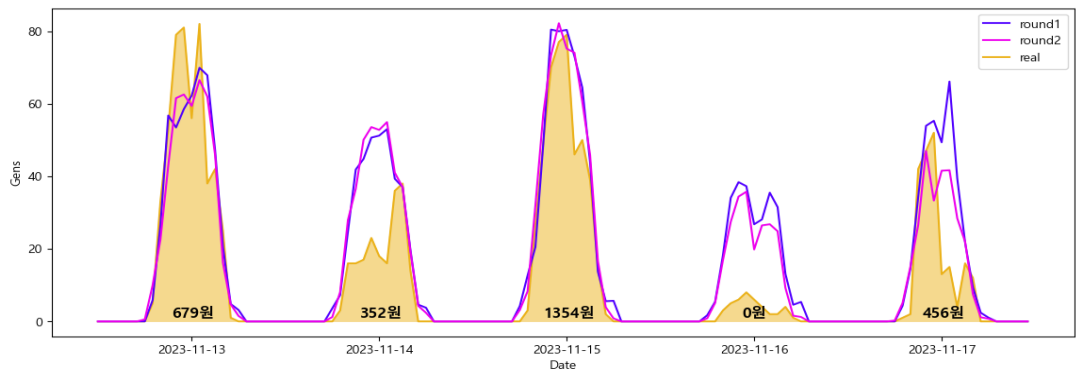
Weight 변화 추이(10시)



Weight 변화 추이(17시)



최종 결과 시각화

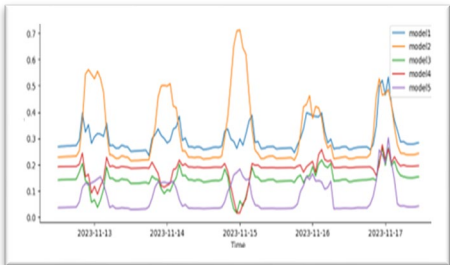


**전체 인센티브 : 2,841원**

\*2023-11-17 17시 입찰만 진행한 경우 : 2,669원

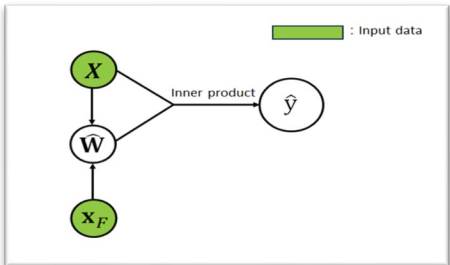
# Conclusion. 결론

## 상황별 앙상블



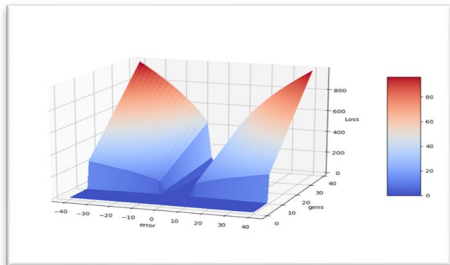
날씨 및 모델 예측값 고려  
→ 가중치 변화

## 데이터 기반 의사결정



데이터와 손실함수 기반 추정  
→ 앙상블 및 개별 모델 확장성

## 시스템 최적화



정량/정성적 손실 최소화  
→ 인센티브 최대화

“시스템 최적해” 를 향한 “앙상블” 의 시작, 仁義禮智信