

# 2024 날씨 빅데이터 콘테스트

## -대회 설명회-

### 과제 1

강수 분야

수치모델 앙상블을 활용한 강수량 예측



1

공모분야 소개

2

분석 데이터 소개  
-기상 데이터

3

날씨 마루 소개

4

검증방법

5

대회 운영 안내

NEXT CONTENTS -----



# 1. 공모분야 소개

## 분석 주제

수치모델 앙상블을 활용한 강수량 예측

-수치수치모델(3시간 단위) 앙상블 강수 확률 자료를 활용하여 (5~9월)의 누적 강수량(계급구간) 예측(단일모델 개발)

## 대상

일반 국민 누구나

- 개인 또는 팀 (구성 인원 10명 이하)

## 분석 데이터

날씨마루(<https://bd.kma.go.kr/>) 분석환경에서 다운

수치모델 데이터: 전국 20개 지점

-예측자료: 수치 앙상블 계급별 강수확률  
: 3시간 주기 자료 (일 2회 (3시간~240시간))

-관측자료: 강수량 (관측값)  
: 3시간 단위 누적 강수량

-학습 데이터 기간: A년 ~ C년의 각 5~9월  
(총 3년중에 5~9월 데이터)

-검증 데이터 기간: D년 5월 ~ 9월  
(특정 5지점의 확률자료)

## ★ 공모범위 ★

- 앙상블 확률 자료를 활용하여 3시간 누적 강수량의 계급구간 예측
- 10개의 강수계급(mm)을 참여자가 반영하여 모델 구성 및 예측 수행
- 모델에 활용된 분석기법 및 방법(제공된 수치모델 데이터만 사용하여 분석)

# 1. 공모분야 소개

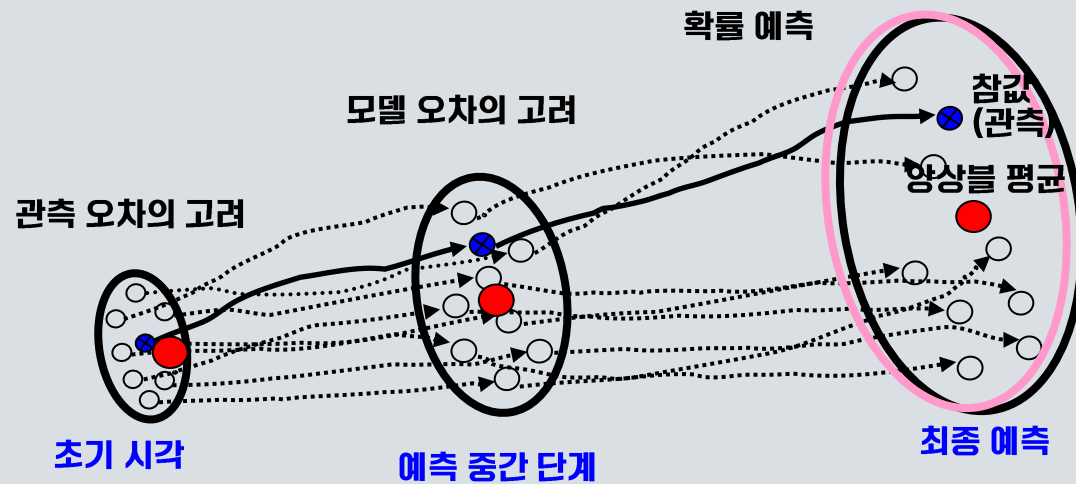
## | 수치예보와 앙상블 예측

수치예보란?

대기의 운동을 설명하는 여러 가지 방정식들을 슈퍼컴퓨터로 풀어 미래의 날씨를 예측하는 과정을 말하며, **수치예보를 수행하기 위해 작성된 프로그램**을 수치예보모델이라고 한다.

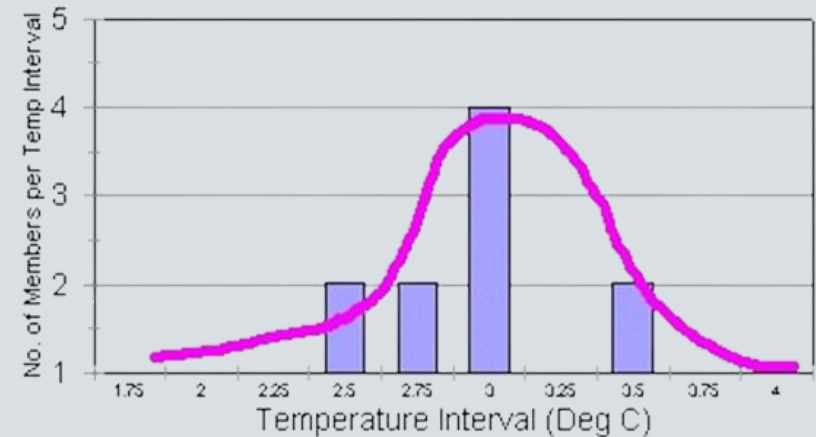
앙상블 예측

날씨 현상이 너무 복잡하고 수치예보모델이 사용하는 방정식들도 완전하지 않기 때문에 100% 완벽한 모델은 존재하기 어렵다. 때문에 하나의 수치예보모델이 가지는 예측 한계를 보완하기 위해, 미세하게 다른 초기자료로부터 시작한 여러 개의 모델들의 예측 결과를 조합하여 확률적으로 미래를 예측하는 복합적인 예측 시스템이다.



Schematic illustration of concepts in ensemble prediction (Wilks, 1995)

Frequency of Members Per Temp Interval  
00 Hr Fcst - 850 mb Temp (deg C) BOS

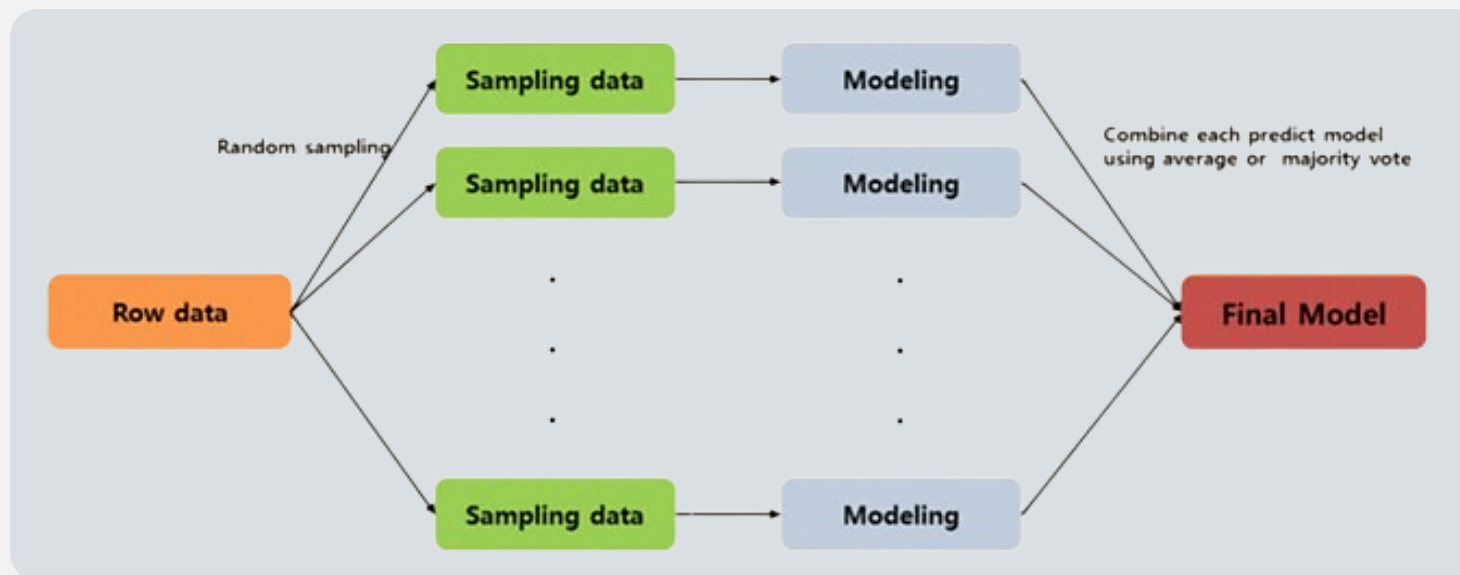


# 1. 공모분야 소개

## | 앙상블 모델

### 앙상블 모델의 특징

- 같은 기간에 대해 여러쌍의 모델을 수행하여 그 결과를 **확률적(통계적)으로 분석**하는 것
- 초기 오차의 확률적 분포가 시간에 따라 어떻게 진화 하는지 분석
- 미래 대기상태에 대한 여러가지 시나리오를 얻을 수 있음
- 한 모델로부터 각기 다른 특성의 초기장을 형성하여 앙상블 멤버를 구성 할 수도 있고, 전혀 다른 독립된 모델들을 묶어 앙상블을 구성할 수도 있음



# 1. 공모분야 소개

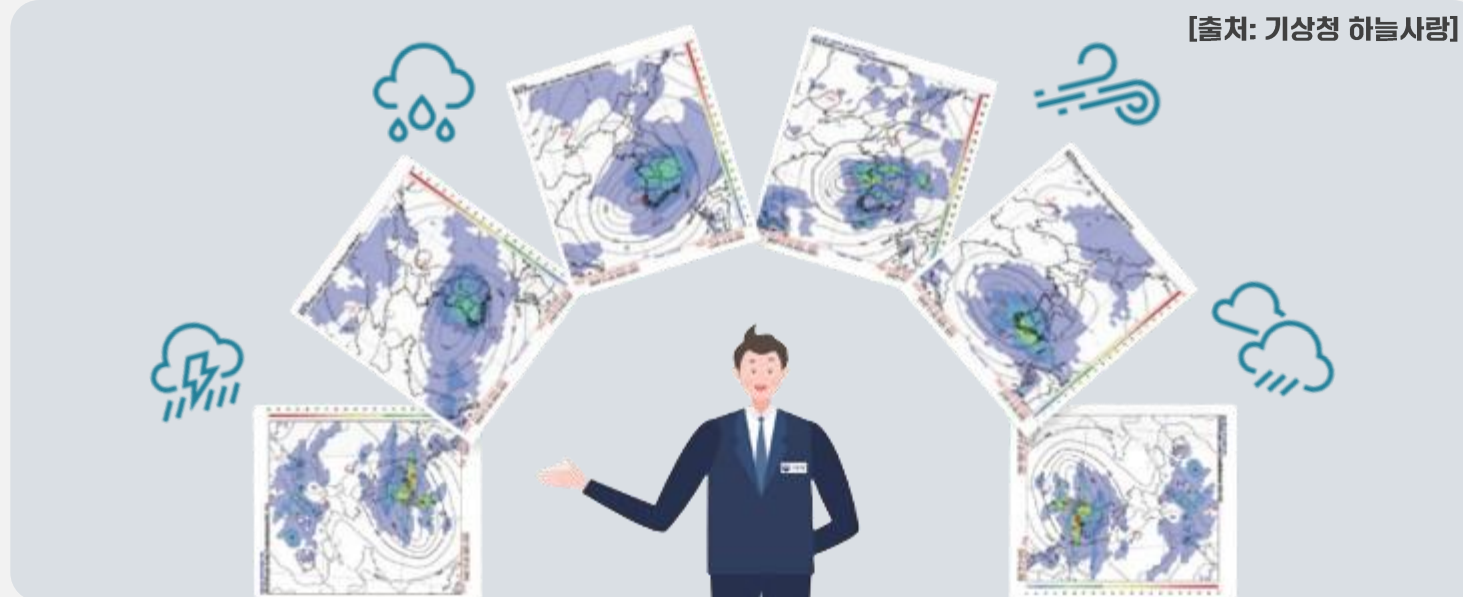
## | 앙상블 모델의 필요성

### 필요한 이유

예를 들어, 앙상블 모델을 20개의 멤버(앙상블 멤버: 앙상블 모델을 구성하는 각각의 수치예보모델들)로 구성하였을 때, 20개의 멤버가 모두 강수를 예측했다면 강수 확률은 100%가 되며 예보관들은 자신감을 가지고 강수를 예측할 수 있다.

반면 20개의 멤버 중 12개만이 강수를 예측했을 시에는 강수가 올 확률과 오지 않을 확률을 고려하여, 강수가 오는 쪽에 조금 더 무게를 두고 예보를 준비할 것이다.

이렇게 앙상블모델을 통해 다양한 시나리오와 그에 따른 **확률을 분석**하면 **예보에 대한 더 합리적인 판단**을 내릴 수 있게 되는 것이다.



### | 수치모델 데이터 변수 구성

변수	설명	변수	설명	변수	설명
TM_FC	기준 발표시각	V02	0.2 mm 이상 누적 확률	V07	10.0 mm 이상 누적 확률
TM_EF	예측 시간	V03	0.5 mm이상 누적 확률	V08	20.0 mm이상 누적 확률
DH	기준시각-예측 시간	V04	1.0 mm이상 누적 확률	V09	30.0 mm이상 누적 확률
STN	AWS 지점 코드	V05	2.0 mm이상 누적 확률	VV	실강수량
V01	0.1 mm 이상 누적 확률	V06	5.0 mm이상 누적 확률	class_interval	강수계급

- DH: 기준발표 시각으로부터 240시간까지 3시간 간격으로 80개의 예측자료의 시간을 의미
- V01~V09 : 양상불 모델의 구간별 강수량 누적 확률 값
- VV(실강수량): 지점별 1시간 강수량자료의 3시간 누적 값
- 강수계급 (class\_interval): 10단계의 강수 계급으로 검증파일(test)에서 -999로 표기된 부분은 결측으로 정답파일에서 변경 없이 제출



## 2. 분석 데이터 소개

## 데이터셋 예시

→ V01~V09: 양상별 구간별 누적확률      → VV: 관측값

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	FC_year	FC_month	FC_day	FC_hour	STN4contest	DH	EF_year	EF_month	EF_day	EF_hour	V01	V02	V03	V04	V05	V06	V07	V08	V09	VV	class_interval
9	A	1	1	9	STN001	174	A	1	8	15	10	10	9	6	5	1	0	0	0	0.0	0
0	A	1	1	9	STN001	177	A	1	8	18	29	28	23	17	12	6	1	0	0	0.0	0
1	A	1	1	9	STN001	180	A	1	8	21	29	28	23	17	12	6	1	0	0	1.3	4
2	A	1	1	9	STN001	183	A	1	9	0	38	34	28	26	14	2	0	0	0	7.3	6
3	A	1	1	9	STN001	186	A	1	9	3	38	34	28	26	14	2	0	0	0	4.8	5
4	A	1	1	9	STN001	189	A	1	9	6	56	52	47	37	25	4	0	0	0	8.0	6
5	A	1	1	9	STN001	192	A	1	9	9	56	52	47	37	25	4	0	0	0	10.7	7
6	A	1	1	9	STN001	195	A	1	9	12	56	49	42	35	23	8	2	0	0	12.6	7
7	A	1	1	9	STN001	198	A	1	9	15	56	49	42	35	23	8	2	0	0	1.9	4
8	A	1	1	9	STN001	201	A	1	9	18	50	42	33	26	18	10	0	0	0	1.5	4
9	A	1	1	9	STN001	204	A	1	9	21	50	42	33	26	18	10	0	0	0	2.2	5
0	A	1	1	9	STN001	207	A	1	10	0	40	38	32	27	16	2	0	0	0	3.1	5
1	A	1	1	9	STN001	210	A	1	10	3	40	38	32	27	16	2	0	0	0	1.0	4
2	A	1	1	9	STN001	213	A	1	10	6	34	32	26	19	15	7	4	0	0	0.1	1
3	A	1	1	9	STN001	216	A	1	10	9	34	32	26	19	15	7	4	0	0	0.1	1
4	A	1	1	9	STN001	219	A	1	10	12	31	28	22	17	13	4	1	0	0	0.0	0

강수계급 (class\_interval): 10단계로 강수량의 계급을 구분함

계급	0	1	2	3	4	5	6	7	8	9
강수량 구간	0.1미만 (무강수)	0.1~0.19	0.2~0.49	0.5~0.99	1.0~1.9	2.0~4.9	5.0~9.9	10.0~19.9	20.0~29.9	30.0이상

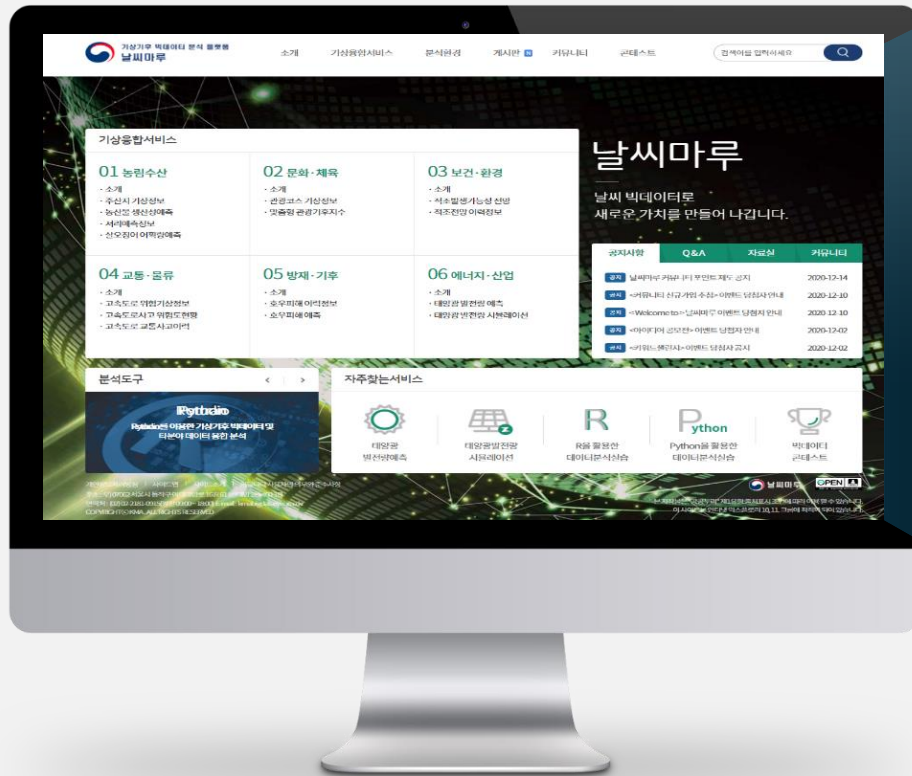
(단위: mm)





### 3. 날씨마루 소개

| 날씨 마루 웹주소: <https://bd.kma.go.kr>



### 3. 날씨마루 소개

#### 기상기후 빅데이터 분석플랫폼

##### 시범서비스

- 빅데이터 기반으로 올바른 날씨 의사결정 지원 <
- 기상기후 빅데이터 융합서비스 개발 <

##### 활용 확산 (R, 파이썬, 포트란)

- 프로그램 기초 교육부터 분석실습 <
- 맞춤형 분석 교육 자료 제공 <



##### 인프라 지원

- > 초기 투자 비용 부담이 높은 스타트업, 기상시장 진입 기업의 기상서비스와 콘텐츠 개발 인프라 지원

##### 분석환경 제공

- > 사업자 혹은 개인이 가진 데이터와 융합할 수 있는 분석 환경 제공

### 3. 날씨마루 소개

기상기후 빅데이터를 누구나! 쉽게! 활용할 수 있어요!  
날씨마루에서는 다양한 기상데이터 다운로드뿐 아니라...

1. R과 Python 분석 환경 제공으로 바로 분석 가능

2. 비정형 분석 도구로 쉽게 워드 클라우드 작성

3. 시각화 도구로 다양한 표로 쉽게 작성

R과 PYTHON



R Programming



Python Programming

워드 클라우드



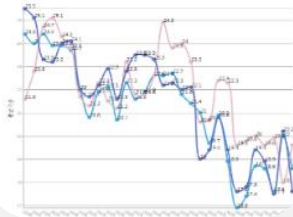
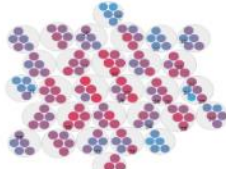
고기온  
작성일 : 날씨마루



9월 지저  
작성일 : 날씨마루

시각화 도구

기온  
작성일 : 김지훈

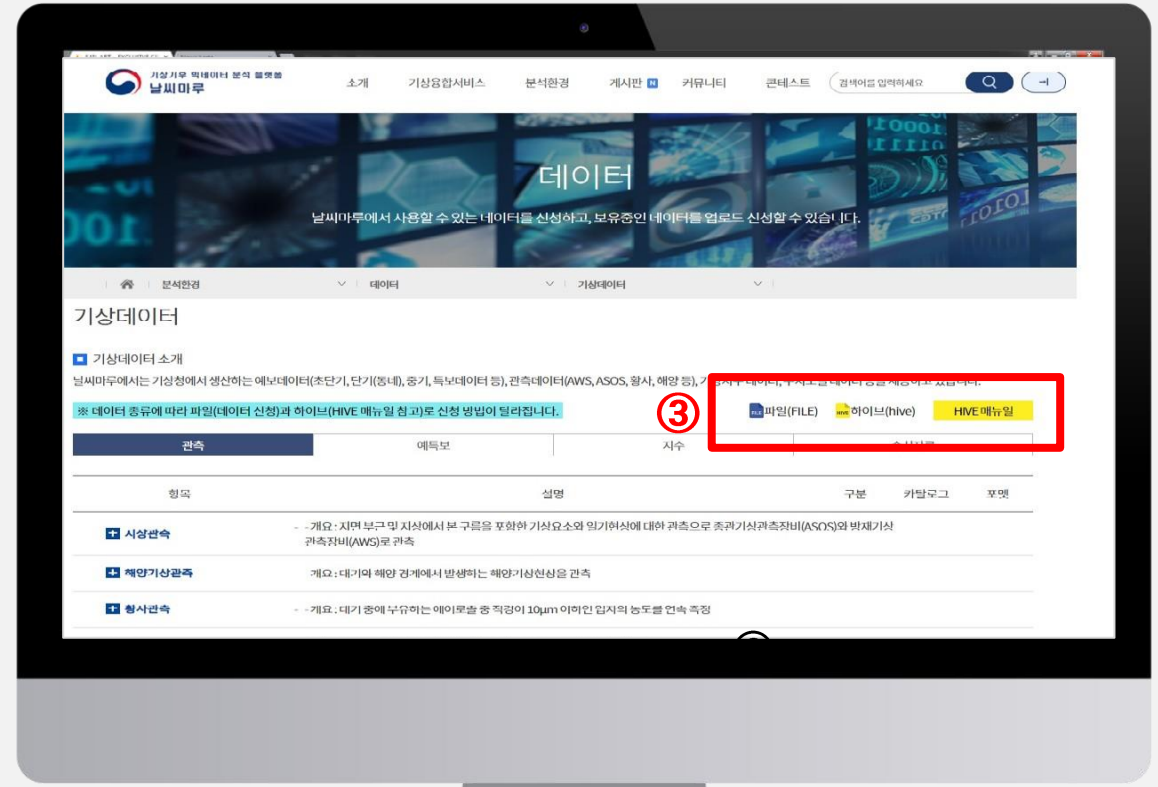


### 3. 날씨마루 소개

#### | 제공 데이터 다운로드 방법



1. 분석 환경 메뉴 접속 -> 신청  
2. 기상데이터 메뉴 접속



3. 하이브(HIVE)  
매뉴얼 접속



### 3. 날씨마루 소개

#### | 기상기후데이터 빅데이터의 활용



**방재**  
지역 맞춤형 호우·대설 피해 예측 서비스



**에너지**  
기상재해에 따른 전력설비 피해 분석 및 예측



**환경**  
기상기후 빅데이터를 이용한 한국연안 적조의  
원인종과 밀도 변동 예측 연구



**농수축산**  
기상기후 민감 작물의 주산지  
기상정보 서비스·생산성 분석



**교통물류**  
날씨에 따른 도로위험도 분석·예측 및 IoT(사물인터넷)  
기반 도로위험 기상정보 생산 기술 개발



**보건건강**

기상기후 빅데이터를 이용한 고혈압 환자의  
응급 합병증 발병 위험 예측 서비스



**산업**

신용카드 정보와 날씨 데이터를 활용한 지역별  
매출 분석 서비스 및 전국 태양광 발전량 예측



**체육관광**

관광코스별 상세 기상정보 제공 서비스 및 메가 스포츠  
이벤트 지원을 위한 기상기후 정보 서비스



**수자원**

가뭄정보의 통합 및 공동활용 기술



## 4. 검증방법

### | 모델 검증 방법

- 날씨 빅데이터 콘테스트 홈페이지([bd.kma.go.kr/contest](http://bd.kma.go.kr/contest)) > 데이터/분석환경 > 검증
- 검증기간: D년 5월~9월(특정 5개 지점)
- 마지막 계산된 모델 평가 점수를 최종 공모안(분석결과서)에 기재하여 제출
  - ※ 검증 결과와 함께 표출되는 검증 순위는 최종 결과가 아님
- 어뷰징 방지를 위해 참가접수 번호 당 검증은 일 2회로 횟수 제한

### 모델 평가 지표

다중 CSI(Critical Success Index: 임계 성공 지수)

$$CSI = \frac{H}{H + M + F}$$

# 4. 검증방법

## | 다중 CSI 계산방법

$$CSI = \frac{H}{H+F+M}$$

H(Hit): H로 시작하는 모든 셀값의 합  
F(Fasle Alarm):F로 시작하는 모든 셀값의 합  
M(Miss): M으로 시작하는 모든 셀값의 합  
C(Correct negative): 무강수로 예측하여 맞춘 경우

X: 강수량(mm)

		예측(mm)									
		X < 0.1	0.1≤x<0.2	0.2≤x<0.5	0.5≤x<1	1≤x<2	2≤x<5	5≤x<10	10≤ x<20	20≤x<30	X≥30
관측 (mm)	X < 0.1	C11	F12	F13	F14	F15	F16	F17	F18	F19	F110
	0.1≤x<0.2	M21	H22	F23	F24	F25	F26	F27	F28	F29	F210
	0.2≤x<0.5	M31	F32	H33	F34	F35	F36	F37	F38	F39	F40
	0.5≤x<1	M41	F42	F43	H44	F45	F46	F47	F48	F49	F410
	1≤x<2	M51	F52	F53	F54	H55	F56	F57	F58	F59	F510
	2≤x<5	M61	F62	F63	F64	F65	H66	F67	F68	F69	F610
	5≤x<10	M71	F72	F73	F74	F75	F76	H77	F78	F79	F710
	10≤x<20	M81	F82	F83	F84	F85	F86	F87	H88	F89	F810
	20≤x<30	M91	F92	F93	F94	F95	F96	F97	F98	H99	F910
	X≥30	M101	F102	F103	F104	F105	F106	F107	F108	F109	H1010



- **시상내역:** 환경부장관상(1팀, 300만원), 기상청장상(1팀, 150만원), 한국기상산업기술원장상(1팀, 50만원), 입선(7팀, 10만원/상장 없음)
- **수상자 지원**
  - (행정 안전부) 공공데이터 활용 창업경진대회 참가 추천
  - (기상청) 차년도 기상기업성장지원센터 입주 지원 시 가점 부여





# 2024 날씨 빅데이터 콘테스트

## -대회 설명회-

---

감사합니다.