


모델 짚는 노인 코드설명 PPT

AI프렌즈 시즌1 온도 추정 경진대회

목차

CONTENTS

- 
- 01 라이브러리 및 데이터
 - 02 데이터 전처리
 - 03 탐색적 자료 분석(EDA)
 - 04 변수 선택 및 모델 구축
- 
- 05 모델 학습 및 검증 (ML)
 - 06 모델 학습 및 검증 (DL)
 - A. Pre-trained Model
 - B. Transfer Model
 - 07 결과 및 결론

01 라이브러리 및 데이터

Library

```
import  
sys, warnings, re, os,  
datetime
```

```
import  
matplotlib, seaborn,  
Ipython.display
```

```
import  
pandas, numpy
```

```
import  
gmean, sp randint, uniform,  
random
```

```
import  
sklearn, lgbm, xgboost,  
pickle
```

```
import  
tensorflow(version: 2.1.0),  
tensorflow.keras
```

Data

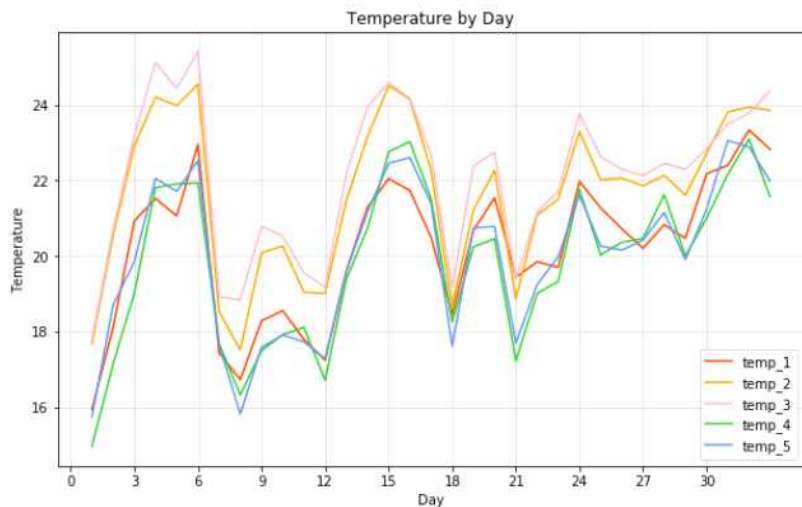
```
train = data/train.csv  
test  = data/test.csv
```

02 데이터 전처리

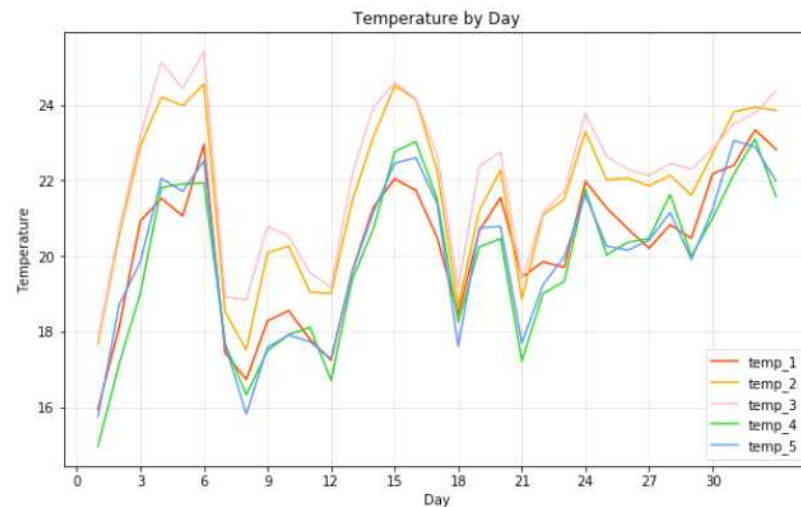
칼럼 설명	기존 칼럼명	변경 칼럼명
교유 번호 (시간순서)	id	id
기온	X00, X07, X28, X31, X32	Temp_1~5
현지기압	X01, X06, X22, X27, X29	LocPr_1~5
풍속	X02, X03, X18, X24, X26	WinSp_1~5
일일 누적강수량	X04, X10, X21, X36, X39	Rain_1~5
해면기압	X05, X08, X09, X23, X33	SeaPr_1~5
일정 누적일사량	X11, X14, X16, X19, X34	Insol_1~5
습도	X12, X20, X30, X37, X38	Humid_1~5
풍향	X13, X15, X17, X25, X35	WinDi_1~5
센서 측정 온도	Y00 ~ Y18	Y00 ~ Y18

03 탐색적 데이터 분석(EDA)

일별 기온 흐름



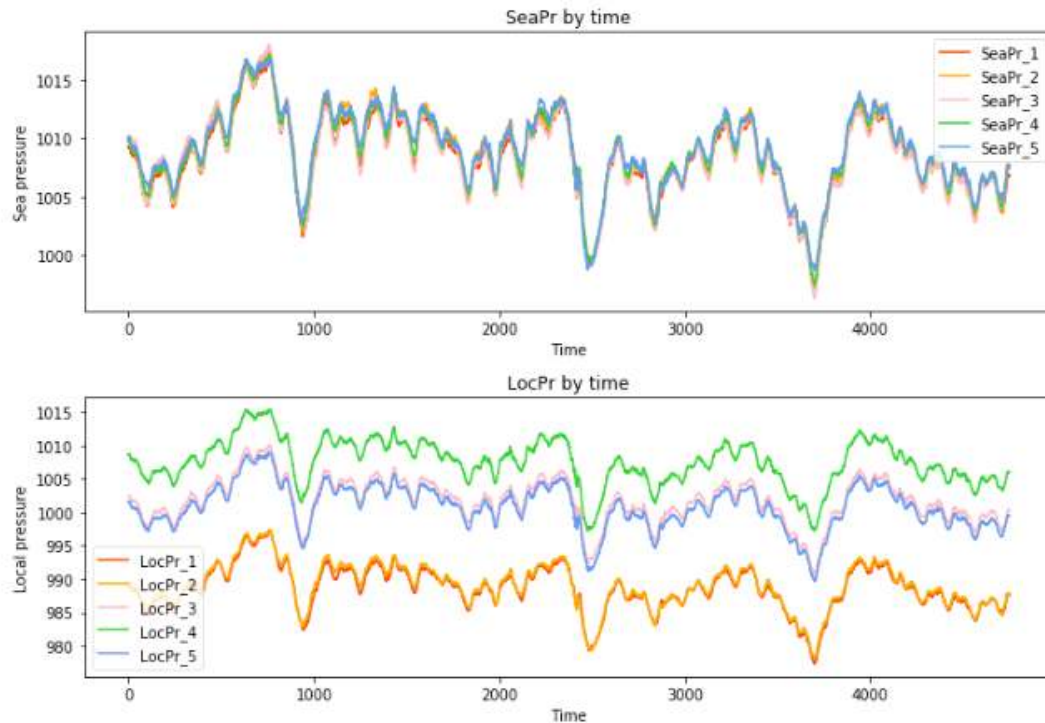
시간별 기온 흐름



시간대, 일별로 5곳의 기온의 차이는 있으나 전체적인 흐름은 유사함을 확인할 수 있다.

03 탐색적 데이터 분석(EDA)

해면기압 & 지면기압

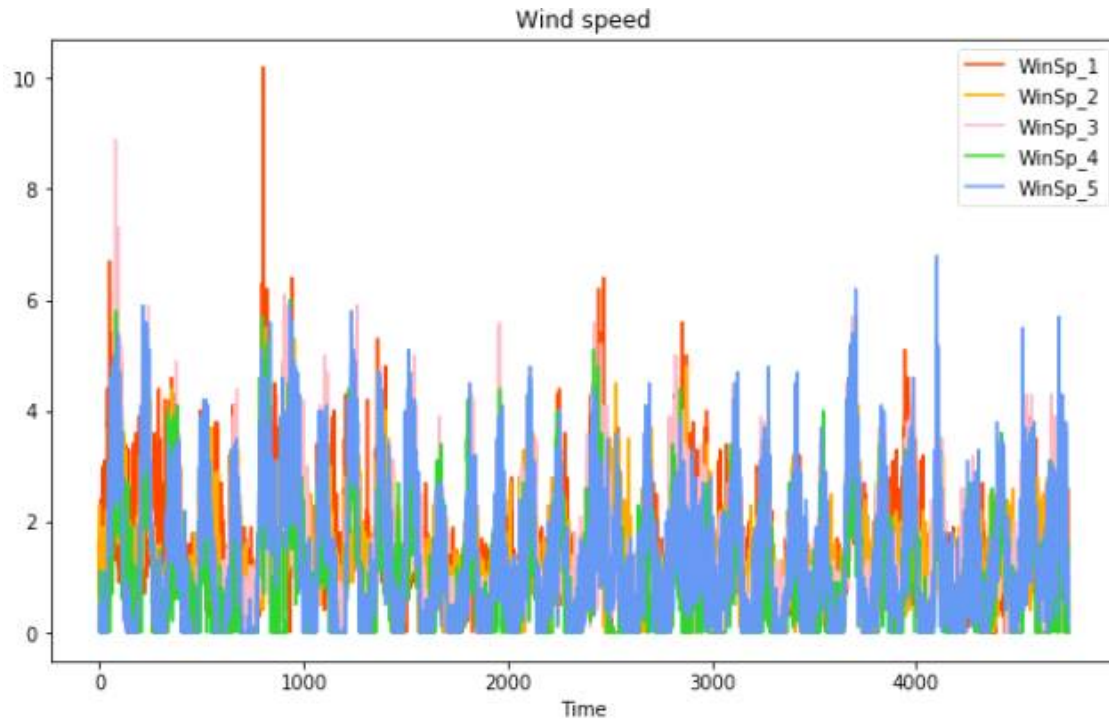


해면기압과 지면기압은 유사한 흐름을 보이며, 순간의 변화를 적은 편이다.
역시 전반적인 흐름은 유사하다.

➡ 기온에 직접적인 영향을 끼친다고 판단하기에는 무리가 있다.

03 탐색적 데이터 분석(EDA)

풍속

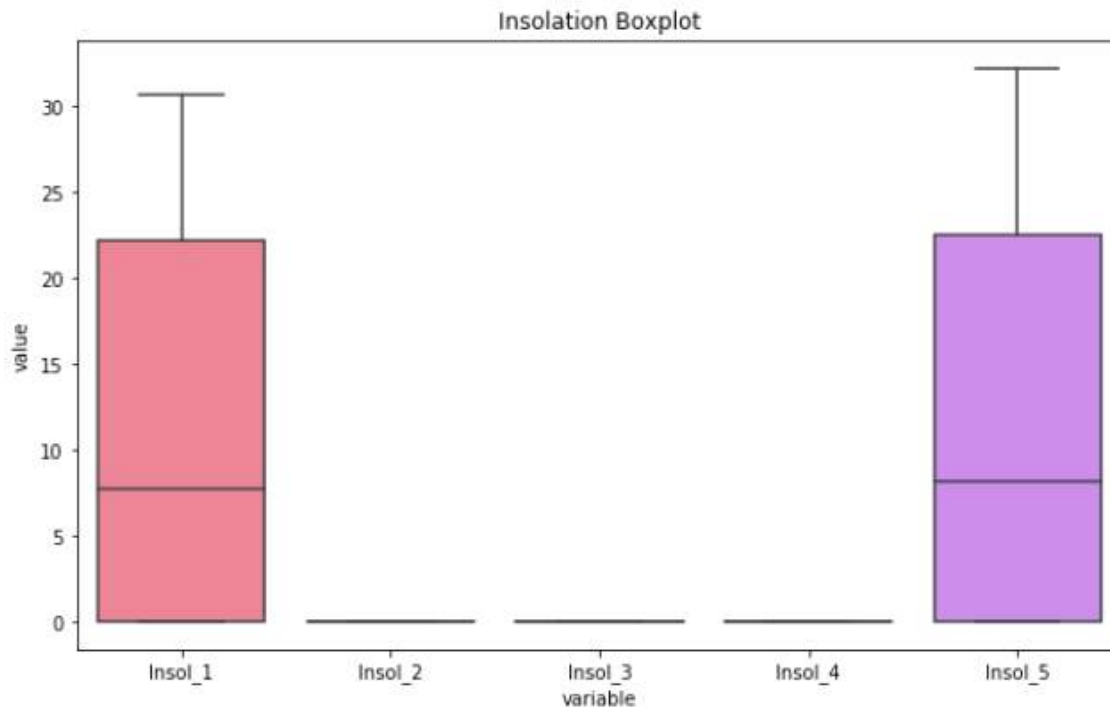


풍속이 날씨의 흐름과 맑음의 정도에 따라서 기온에 미치는 영향이 달라지는데, 그 정도를 파악하는데 문제가 있다.

➡ 변수에 활용하기에는 어려움이 있음을 확인하였다.

03 탐색적 데이터 분석(EDA)

일사랑

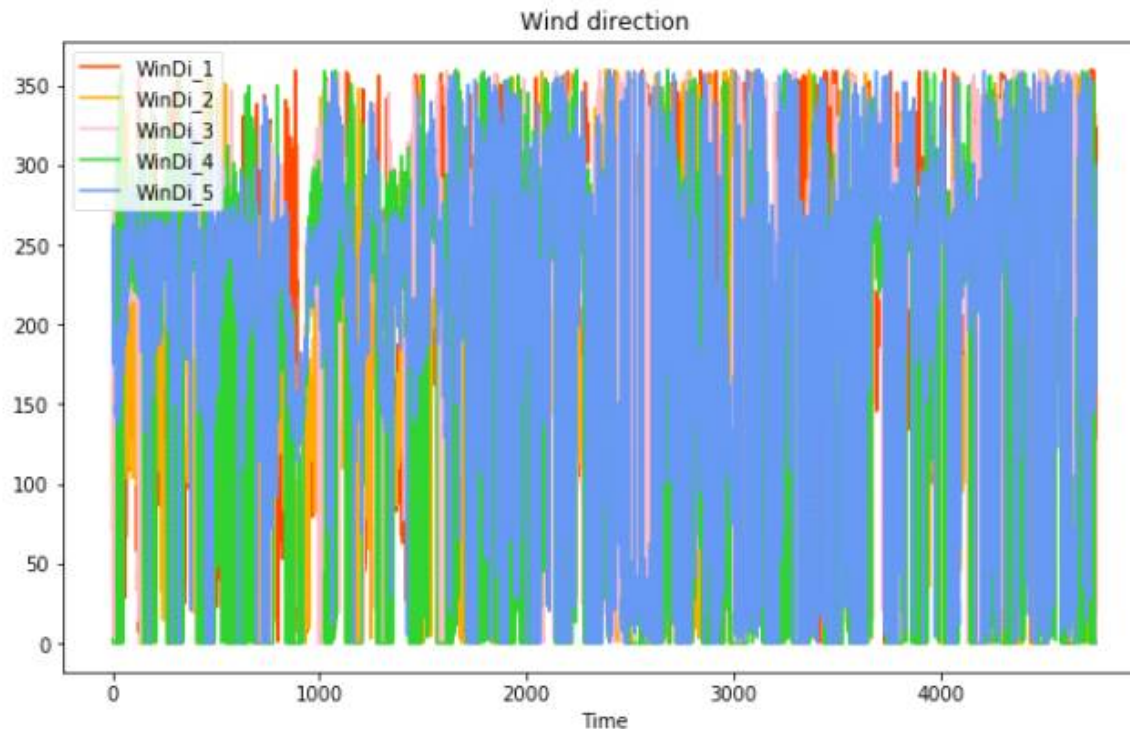


Insolation 1,5를 제외한 다른 피쳐는 모든 데이터가 '0'이다.

➡ Insolation 2-4 피쳐는 삭제한다.

03 탐색적 데이터 분석(EDA)

풍향

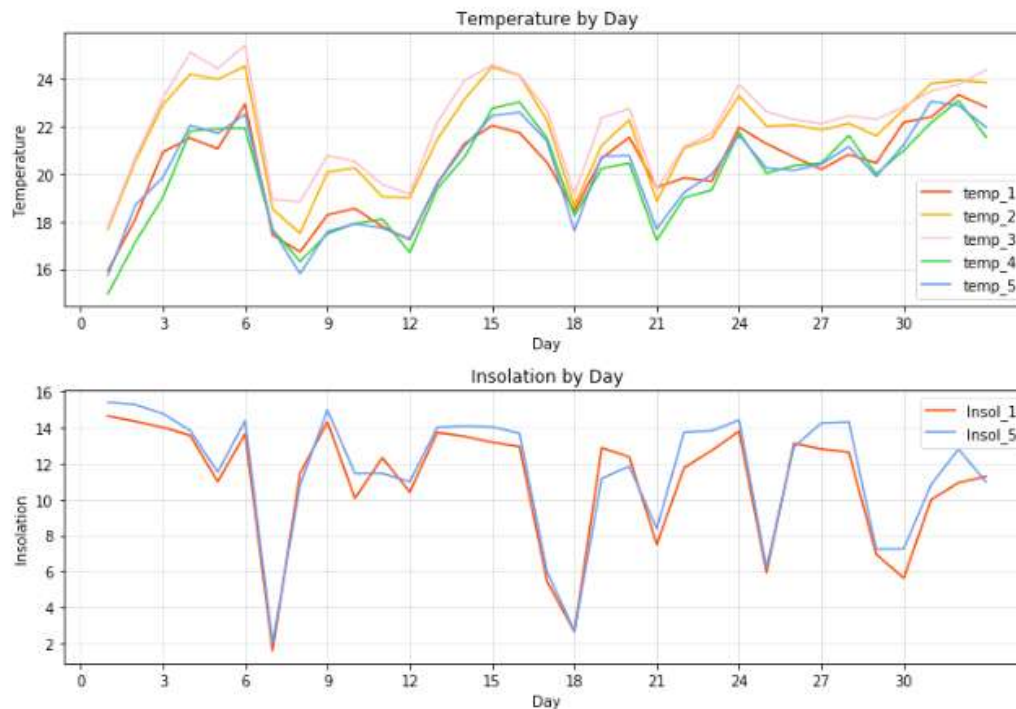


풍향은 주어진 데이터로서의 의미보다는 **방향으로서의 의미가 중요하다고** 판단하였다.

➡ 45도씩 방향을 구분하여 원핫인코딩을 진행하였다.

03 탐색적 데이터 분석(EDA)

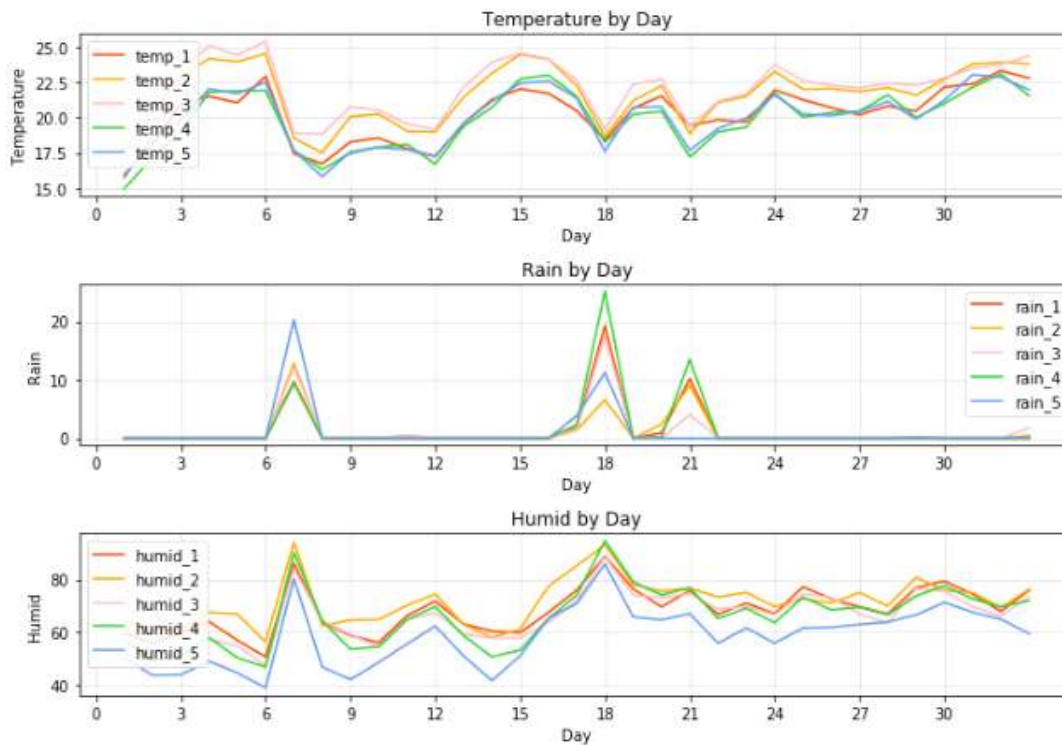
기온과 일사량 비교



기온이 낮은 날에는 일사량이 낮으며, 기온이 높은 날에는 일사량이 높은 흐름을 확인할 수 있다.

03 탐색적 데이터 분석(EDA)

기온, 비, 습도 비교



기온이 낮아지는 시점이 비가 온 시점과 일치함을 확인할 수 있으며, 비가 오는 시점에 습도 또한 높음을 확인할 수 있다.

03 탐색적 데이터 분석(EDA)

종합

각 변수별 흐름을 시각화한 결과, [일사량, 강수, 습도]는 기온과 높은 상관관계를 확인할 수 있었다. 단, 습도와 강수는 역시 높은 관계를 보인다.

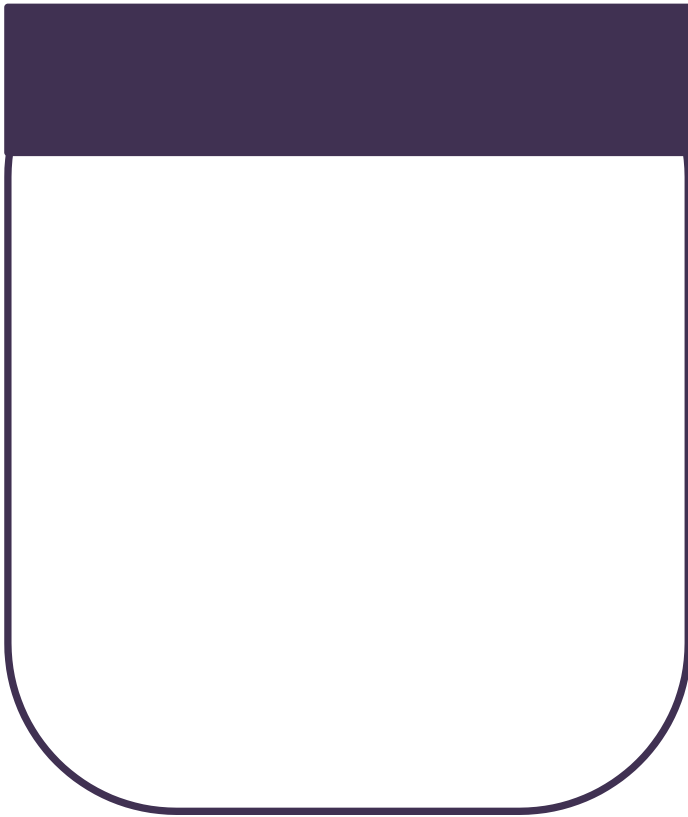
‘일사량’은 기온과 양의 상관관계를 보이며, ‘강수와 습도’는 기온과 음의 상관관계를 보인다.



이를 바탕으로 Feature Engineering 단계에서는 ‘일사량’, ‘강수’, ‘습도’, ‘기온’ 변수를 적극적으로 활용하여, 나머지 지표들은 응용하여 활용하는 방향을 설정하였다.

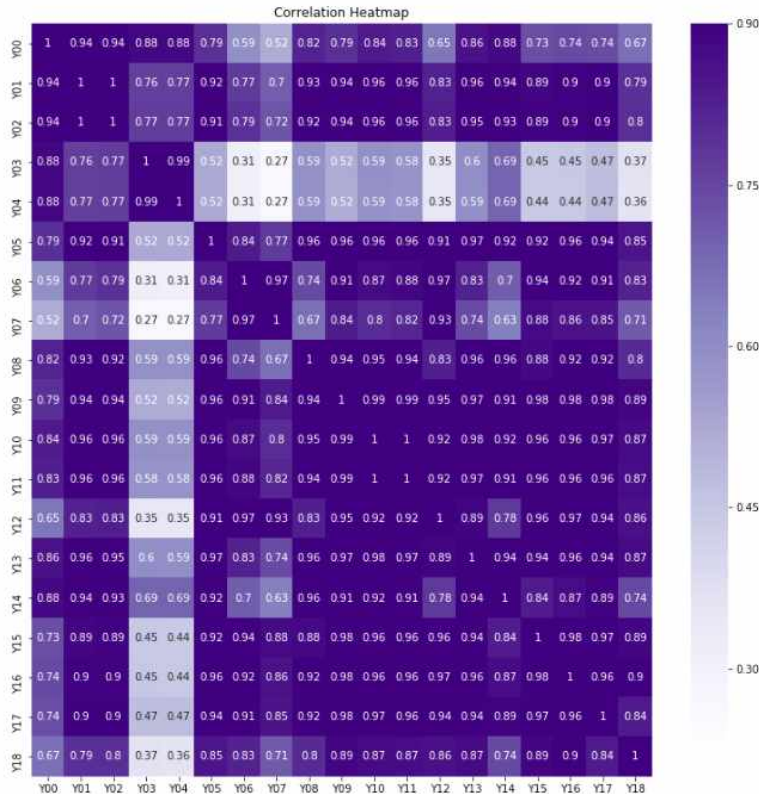
04 변수 선택 및 모델 구축

Features



05 모델 학습 및 검증 (ML)

Train [Y18] 대치



'Y18'과의 상관계수가 **0.83**이상의
Y열 추출

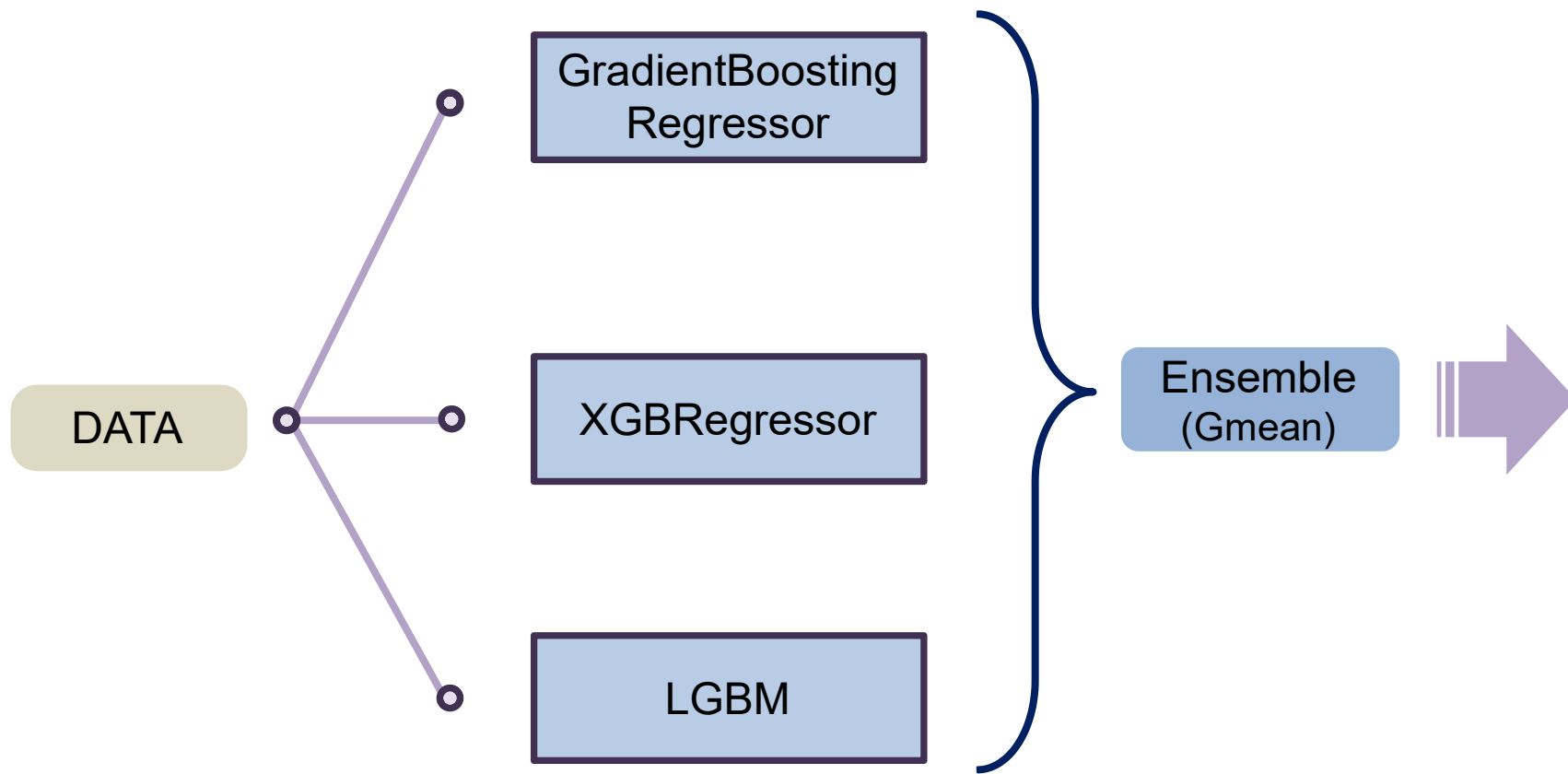
Y =
['Y05','Y06','Y09','Y10','Y11','Y12','Y13','Y15','Y16','Y17']

첫 30일 Y18 =
추출된 Y열들의 평균으로 대체

▲ Y00~Y18 사이의 관계 파악 (By InCastle)

05 모델 학습 및 검증 (ML)

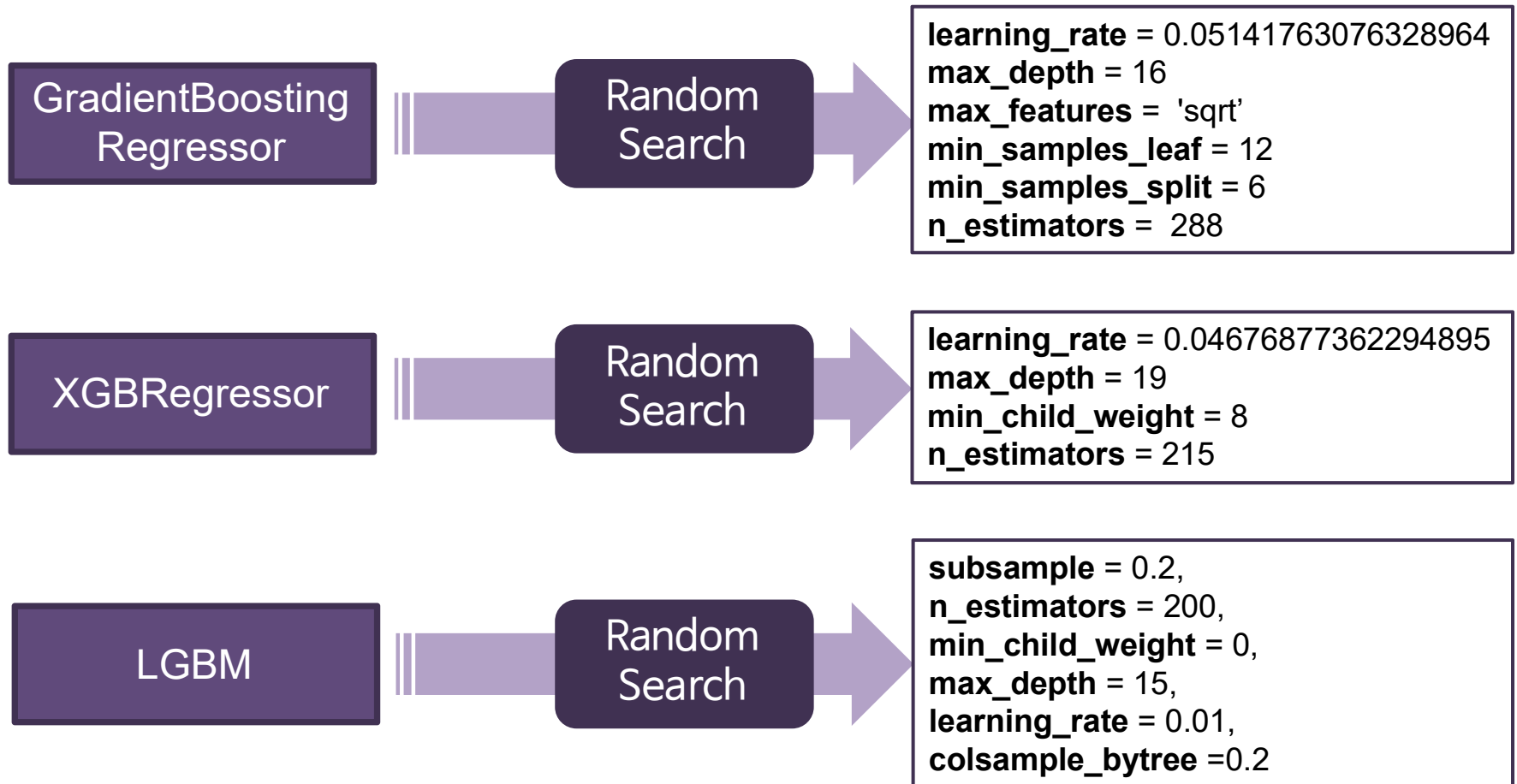
Machine Learning Process



05 모델 학습 및 검증 (ML)

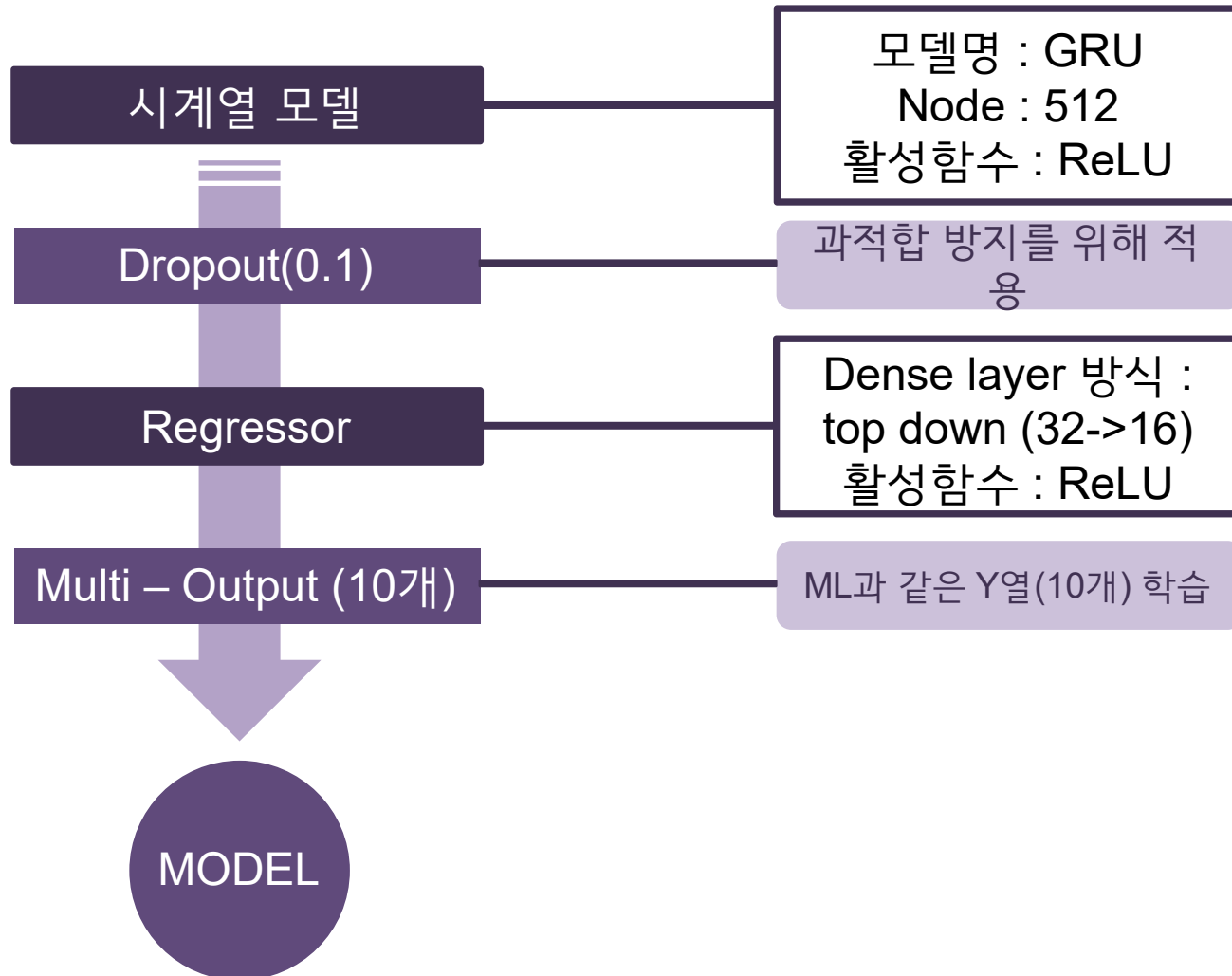
Machine Learning Models

Best Parameters



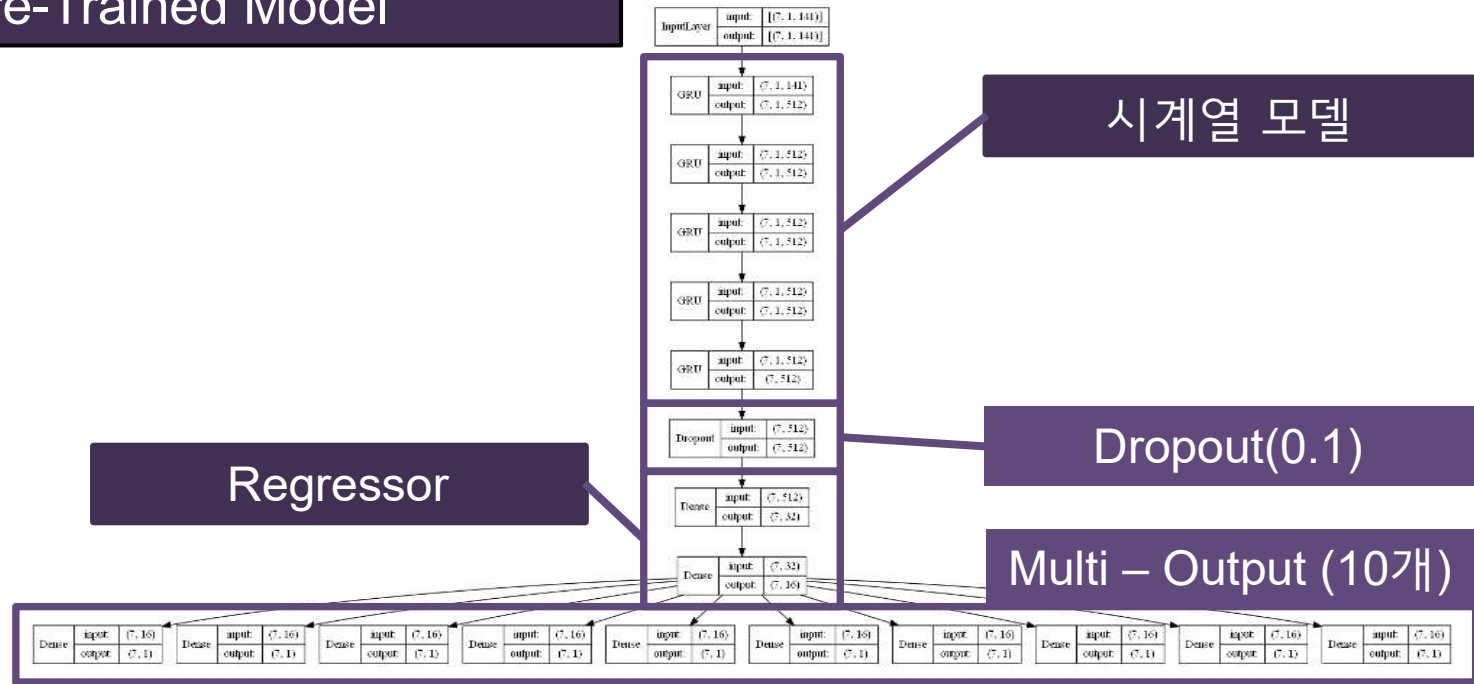
05 모델 학습 및 검증 (DL)

A. Pre-Trained Model



05 모델 학습 및 검증 (DL)

A. Pre-Trained Model



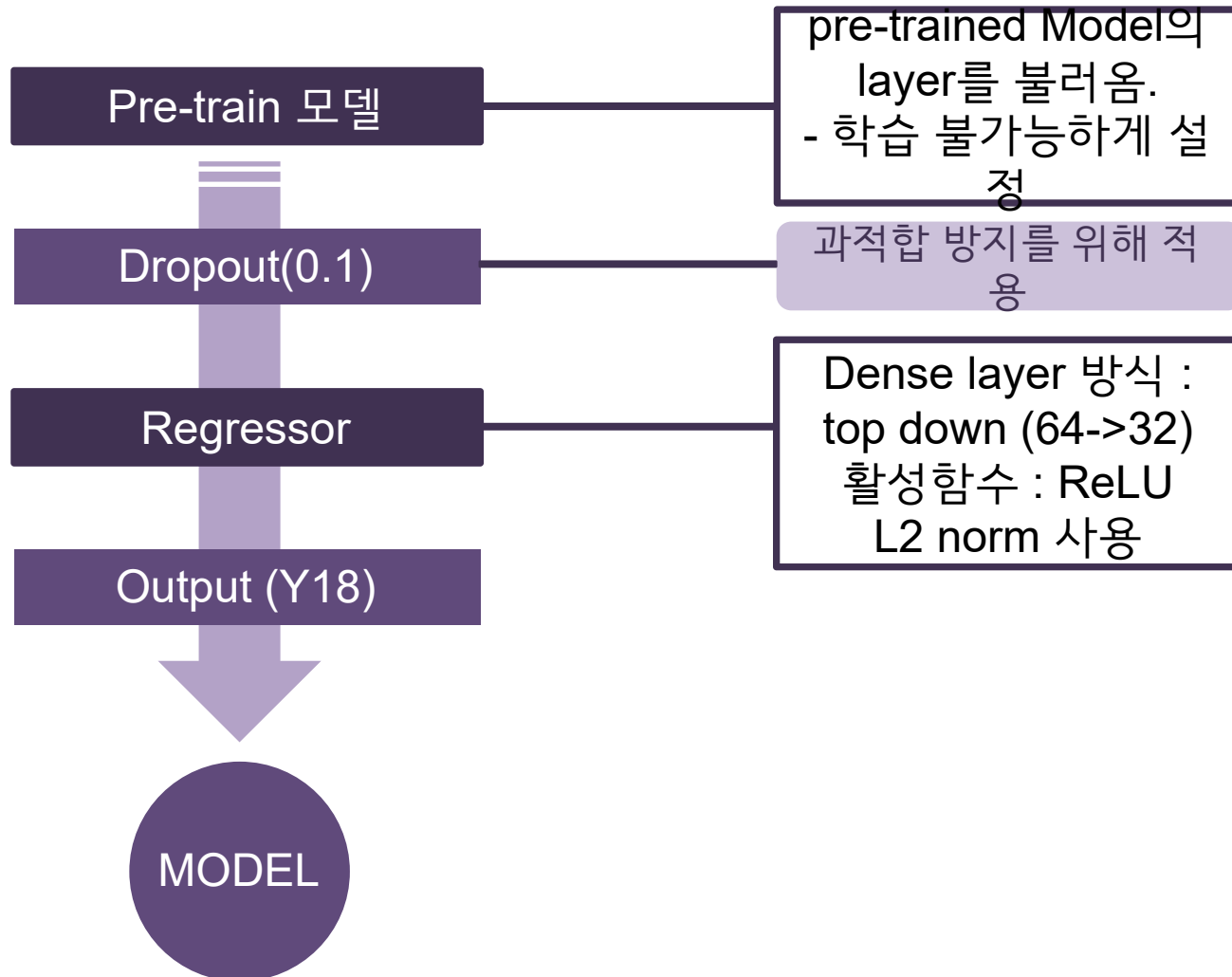
▲Pre-Trained Model 구조도

모델 컴파일
옵티마이저 : Adam
로스: mse
매트릭: mse_AI프렌즈

모델 학습
Epochs : 150
Batch_size : 128

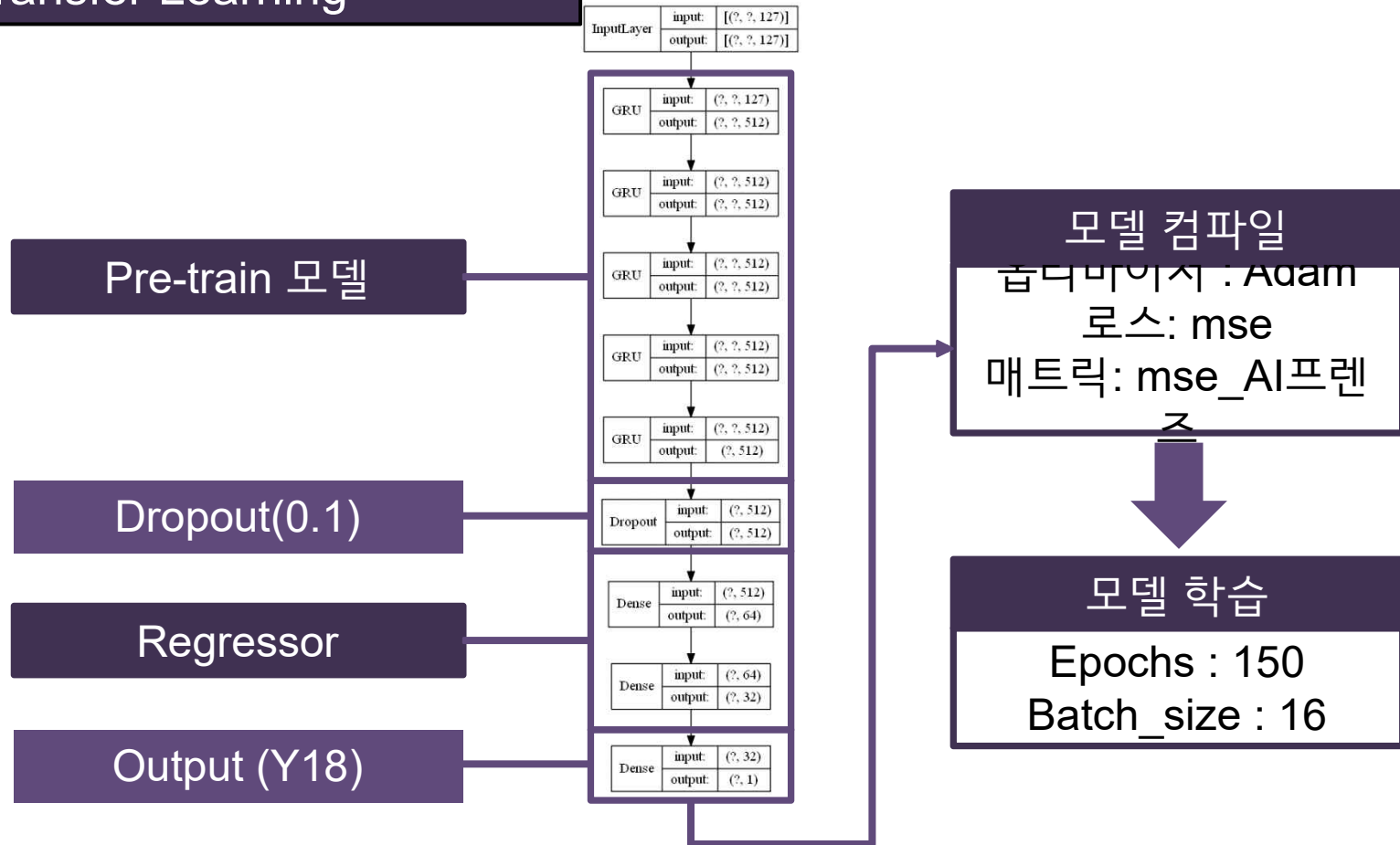
05 모델 학습 및 검증 (DL)

B. Transfer Learning



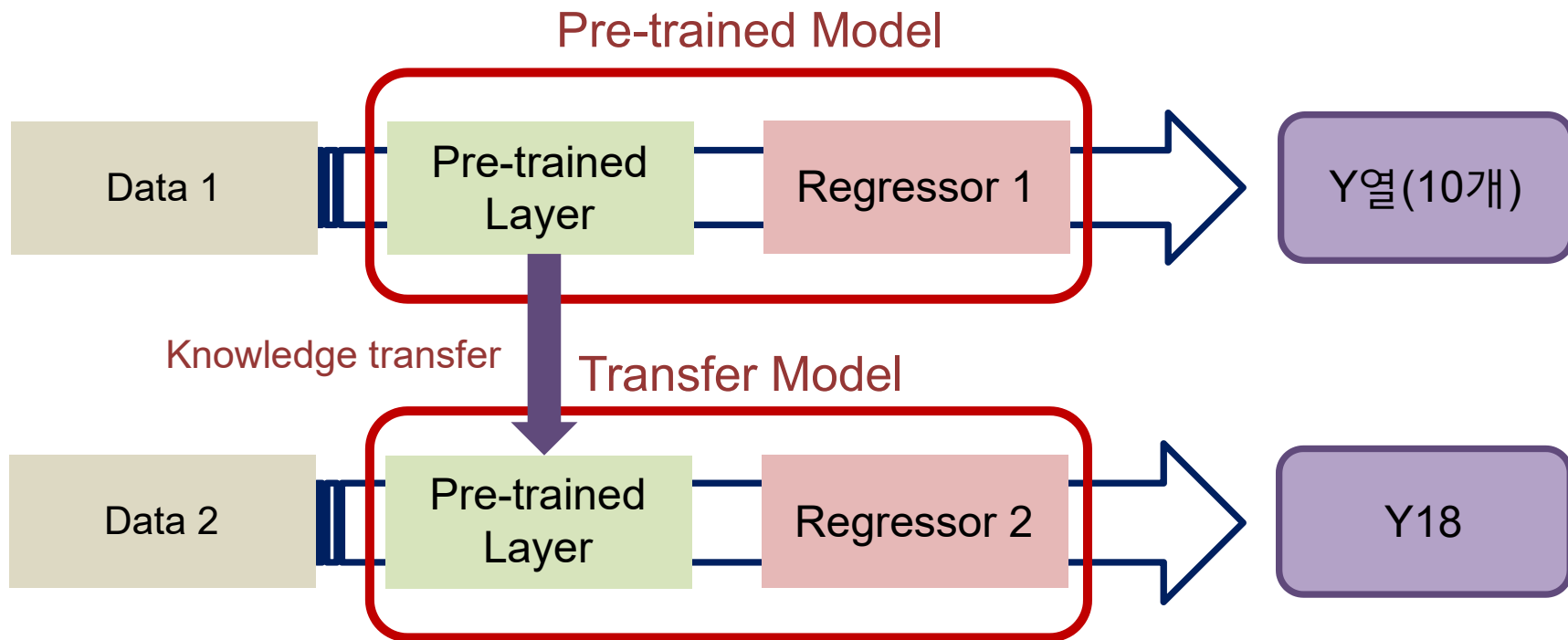
05 모델 학습 및 검증 (DL)

B. Transfer Learning



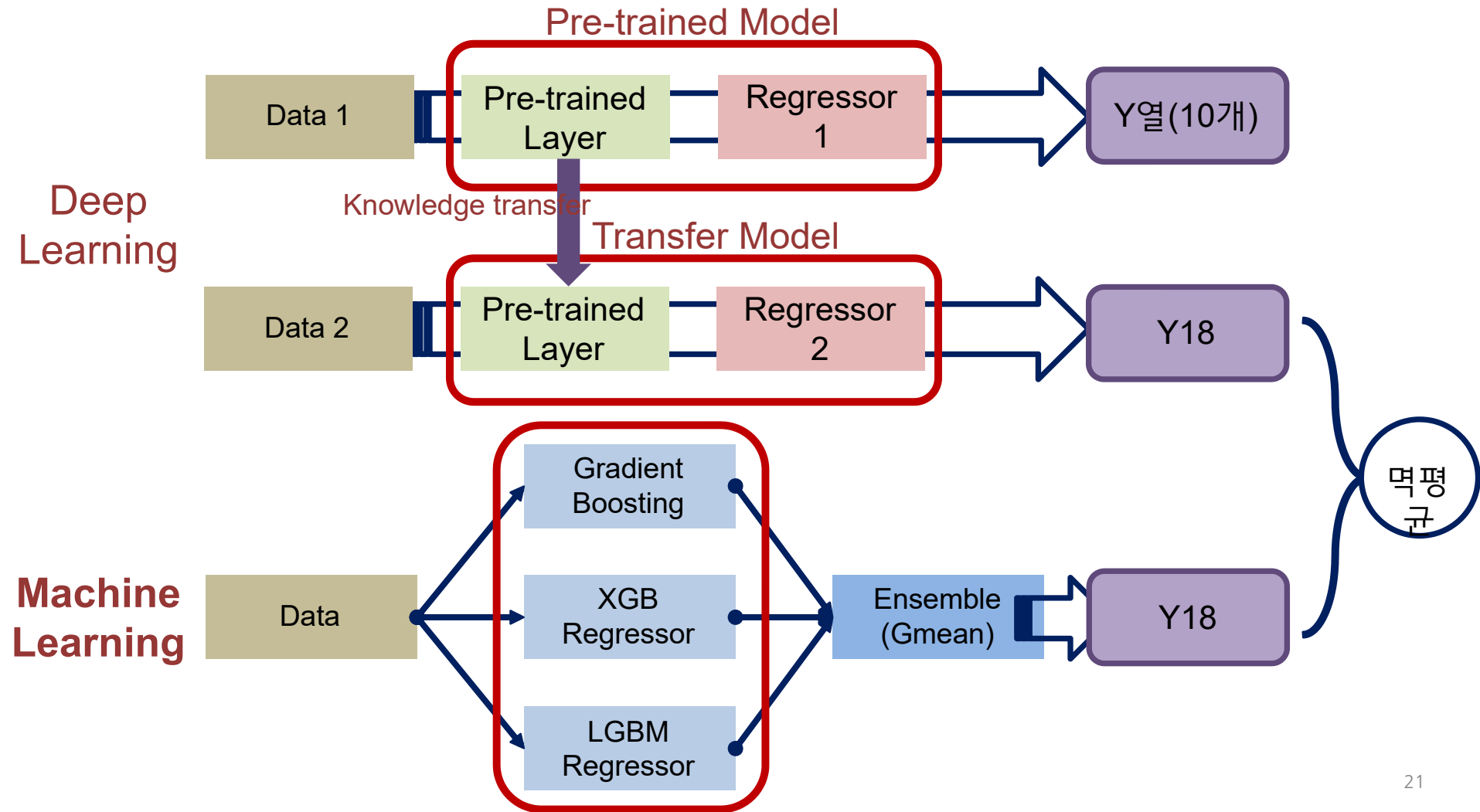
05 모델 학습 및 검증 (DL)

Deep Learning Process



05 모델 학습 및 검증 (ML & DL)

Total Process



감사합니다.
