

# SAI 하반기 70.0 최종 결과 보고

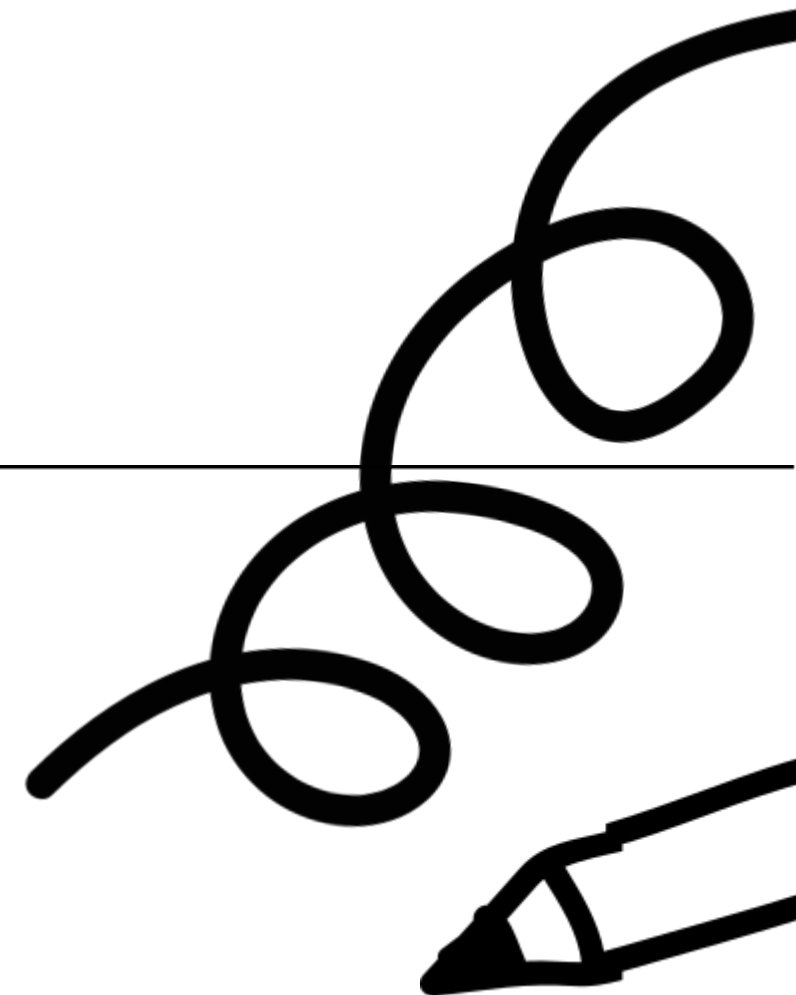


70.0

컴퓨터공학과 21011591 서민정

컴퓨터공학과 19011502 최락원

컴퓨터공학과 19011558 서상민



# *INDEX*

- 학습 주제 및 목표
- 활동 과정
- 학습 성과
- 결론 및 느낀 점
- 앞으로의 계획





# 학습 주제 및 목표

## 학습 주제

- 인공지능 경진대회 플랫폼인 '데이콘'을 활용한 대회 참가 및 실무 능력 향상
- 백엔드 관련 지식 습득 및 데이터/인공지능 분야에서의 활용 방법 탐색

## 학습 목표

- 인공지능을 활용한 데이터 분석 및 예측 경험
- 예측 성능보다는 데이터 분석, 예측 기법에 대한 이론적 내용과 필요성 인식
- Pandas, numpy, scikitlearn 등의 분석 기법과 수학적 이론을 기반으로 폭넓은 분석과 다양한 인사이트 도출
- 백엔드의 이해를 통한 인공지능에 활용되는 데이터 수집 및 가공 방식의 이해

학습 준비(계획서 & Notion)

# 활동 과정

## 1회차

- 세부학습 계획서 작성
- Notion 페이지 공유

전공심화공동체 세부학습계획서

팀 명	70.0	팀 장	서민정
		팀 원	서상민
		팀 원	최락원
		팀 원	
학습주제	인공지능 경진대회 플랫폼인 '데이콘'을 활용해 총 2개의 대회 참가 및 실무 능력 향상		
학습목표	<div>- 실제 비즈니스 문제를 해결하는 분석 경험을 통한 문제해결 능력 강화</div> <div>- 첫 번째 데이콘 경진대회 도전 : HD현대 AI Challenge - 조선 해양 분야 데이터 분석</div> <div>링크 : <a href="https://dacon.io/competitions/official/236158/overview/description">https://dacon.io/competitions/official/236158/overview/description</a></div> <div>- 두 번째 데이콘 경진대회 도전 : 차후 개최된 대회 혹은 적절한 난이도의 데이터를 선정</div>		
활동규칙	<div>- 시간 엄수</div> <div>- 과제 기한 내에 마치고</div> <div>- 역할에 최선을 다해 임하기 (민정 : 사진촬영, 상민 : 공간대여, 락원 : 사무, 일지 : 상민, 민정, 락원 순)</div> <div>- 활동에 최선을 다하기</div>		



## 전공심화공동체 - 70.0

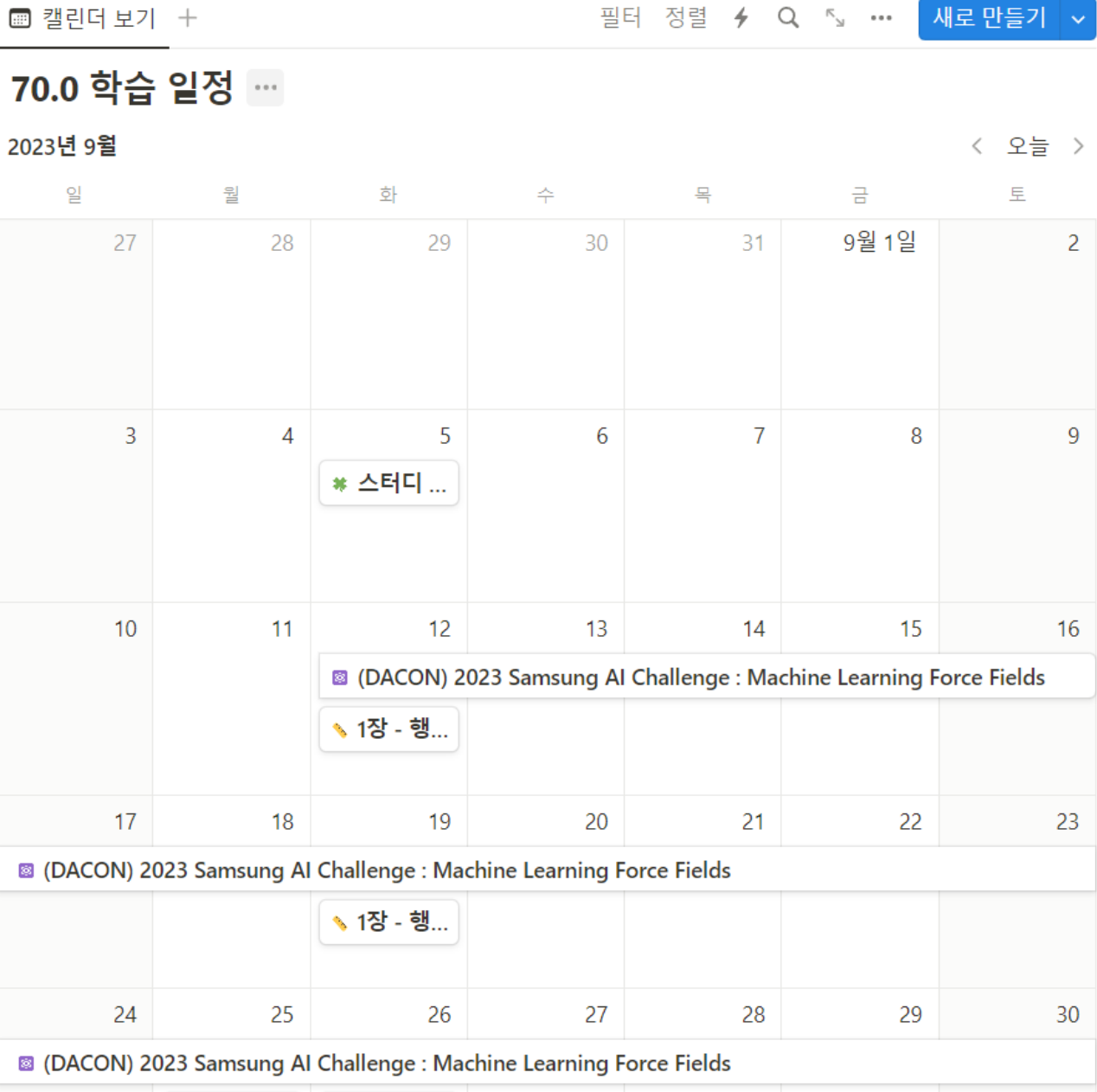
### 활동 페이지

- ☀ 공유 학습 공간
- 💻 개인 학습 공간
- 📅 회의 및 스터디

### 학습 일정

📅 캘린더 보기

## 전공심화공동체 - 70.0

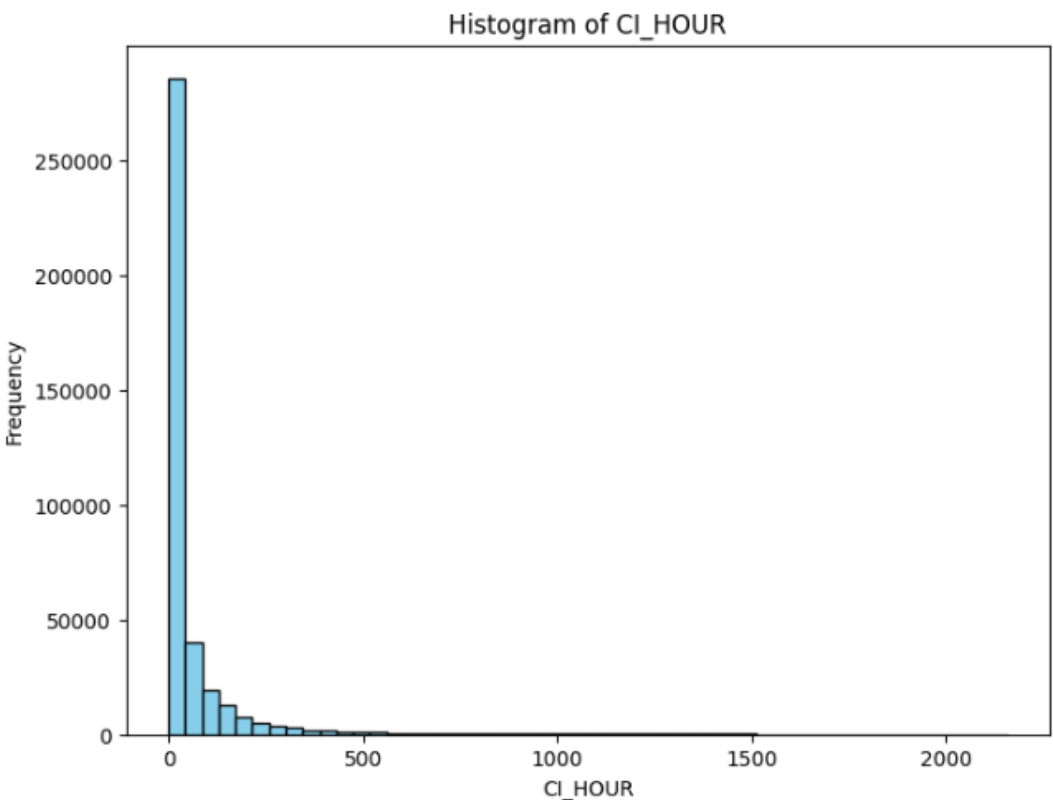


# 활동 과정 ↗

HD현대 AI Challenge (상위 35%)

## 2회차

- HD현대 AI Challenge 참여
- 데이터 내려 받기
- 데이터 개요 파악



## 3회차

- 데이터 전처리 및 시각화, 이상치 탐지
- 독립/종속 변수 간 상관관계 분석
- OLS 분석을 통한 다중 공선성, 정규성 확인
- 머신러닝 모델링

OLS Regression Results			
Dep. Variable:	CI_HOUR	R-squared:	0.068
Model:	OLS	Adj. R-squared:	0.068
Method:	Least Squares	F-statistic:	1101.
Date:	Mon, 30 Oct 2023	Prob (F-statistic):	0.00
Time:	13:08:45	Log-Likelihood:	-2.5571e+06
No. Observations:	391939	AIC:	5.114e+06
Df Residuals:	391912	BIC:	5.115e+06
Df Model:	26		
Covariance Type:	nonrobust		

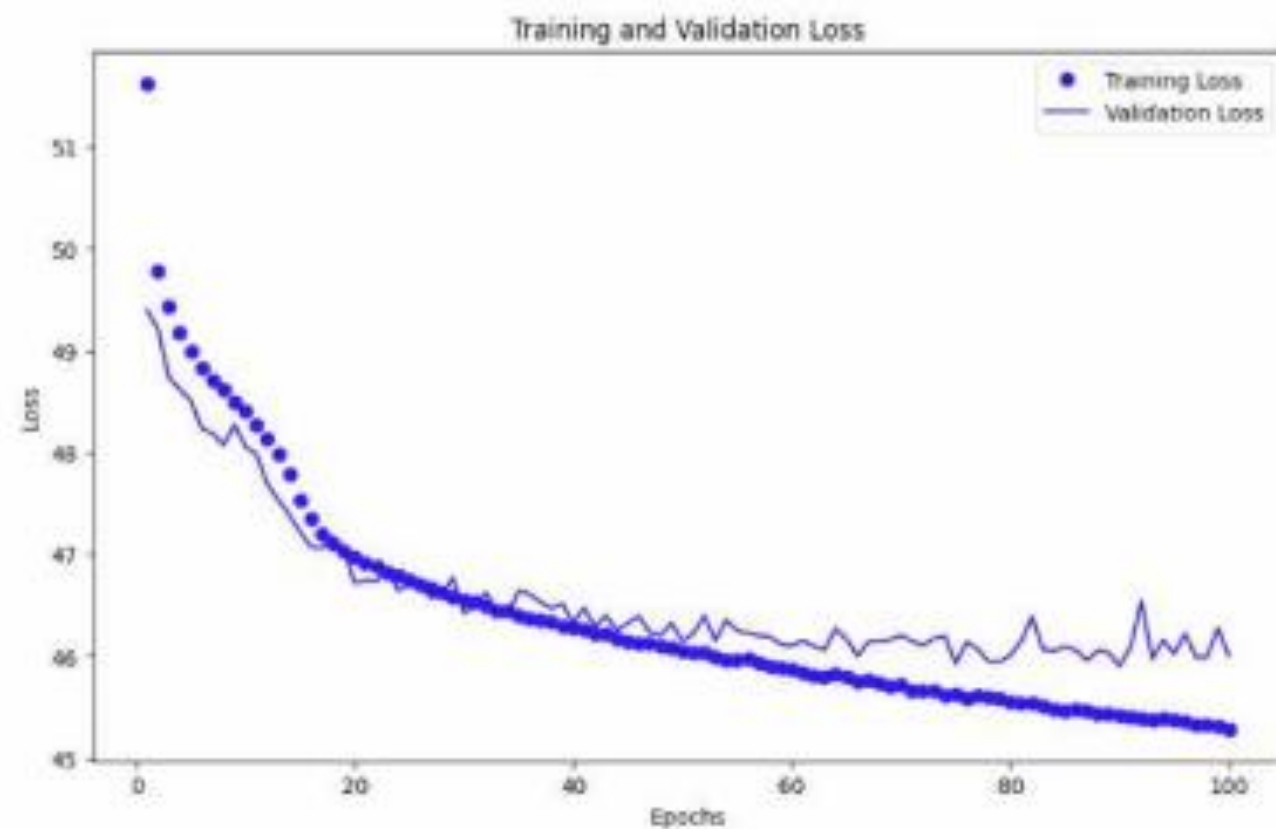
	coef	std err	t	P> t	[0.025	0.975]
const	-3066.4881	238.013	-12.884	0.000	-3532.987	-2599.989
ARI_CO	2.1074	0.049	43.444	0.000	2.012	2.203
ARI_PO	0.0109	0.011	1.019	0.308	-0.010	0.032
SHIP_TYPE_CATEGORY	-19.8656	0.383	-51.851	0.000	-20.617	-19.115
DIST	1.0947	0.009	119.895	0.000	1.077	1.113
ID	-0.0001	3.63e-05	-4.120	0.000	-0.000	-7.84e-05
BREADTH	-0.8530	0.070	-12.204	0.000	-0.990	-0.716
BUILT	0.7860	0.040	19.874	0.000	0.709	0.864
DEADWEIGHT	0.0002	1.57e-05	9.798	0.000	0.000	0.000
DEPTH	1.1759	0.087	13.593	0.000	1.006	1.346
DRAUGHT	-1.6430	0.103	-15.939	0.000	-1.845	-1.441
GT	0.0007	2.88e-05	23.253	0.000	0.001	0.001
LENGTH	-0.2464	0.012	-20.374	0.000	-0.270	-0.223
SHIPMANAGER	7.238e-05	0.000	0.498	0.618	-0.000	0.000

# 활동 과정

HD현대 AI Challenge (상위 35%)

## 4회차

- 데이터 분석 결과 최종 검증
- 데이콘에 결과 제출
- 진행 과정, 성과 리뷰



## 5회차

- 최종 리더보드 순위 확인
- Best Model 분석

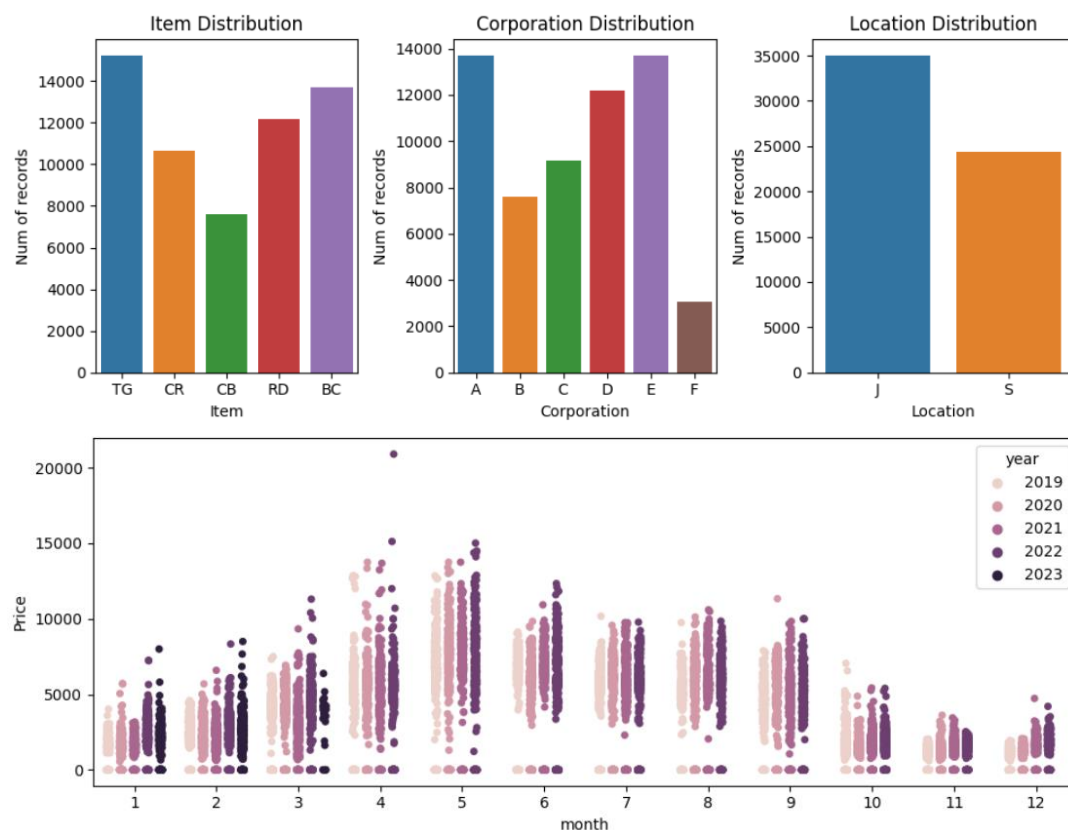
Rank	Score	Submit
117 seomj0	45.90871	24

# 활동 과정

제주 특산물 가격 예측 AI 경진대회 (상위 3%)

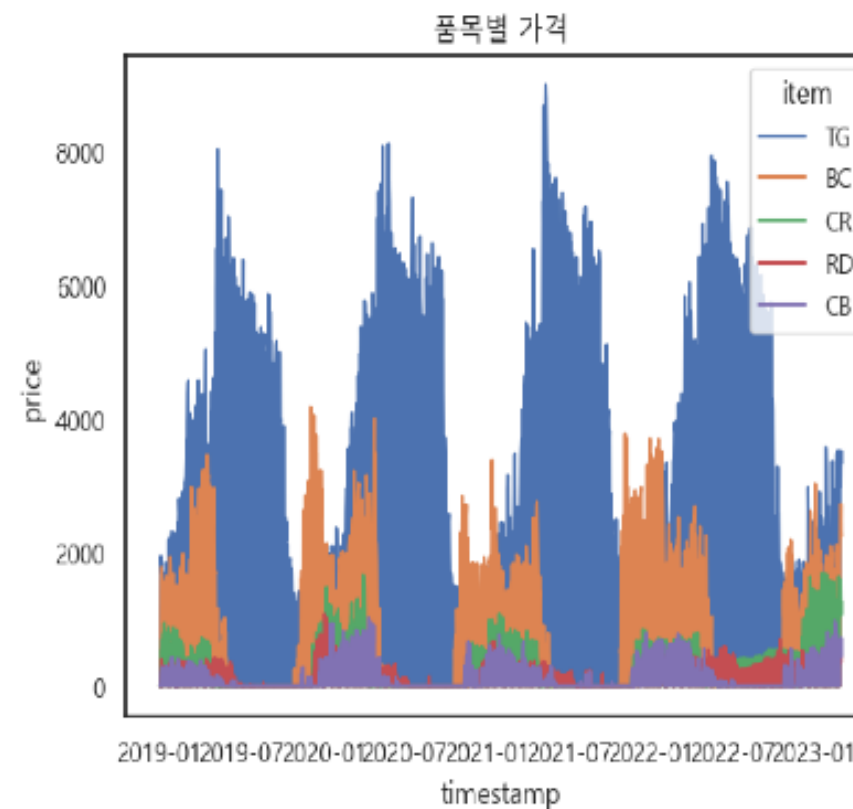
## 6회차

- 제주 특산물 가격 예측 AI 경진대회 참여
- 데이터 내려 받기
- 데이터 개요 파악



## 7회차

- 시계열 데이터 특성에 맞게 시각화
- 각 변수의 전처리 진행
- 다양한 AI 모델을 활용한 모델링



	ID	timestamp	item	corporation	location	supply(kg)
1125	TG_A_J_20220130	2022-01-30	TG	A	J	43455.0
2648	TG_A_S_20220130	2022-01-30	TG	A	S	38439.0
4171	TG_B_J_20220130	2022-01-30	TG	B	J	3294.0
5694	TG_B_S_20220130	2022-01-30	TG	B	S	80659.0
7217	TG_C_J_20220130	2022-01-30	TG	C	J	51458.4
8740	TG_C_S_20220130	2022-01-30	TG	C	S	29426.0
10263	TG_D_J_20220130	2022-01-30	TG	D	J	13505.0
11786	TG_D_S_20220130	2022-01-30	TG	D	S	5825.0
13309	TG_E_J_20220130	2022-01-30	TG	E	J	12705.0
14832	TG_E_S_20220130	2022-01-30	TG	E	S	20073.0
27016	CB_A_J_20220130	2022-01-30	CB	A	J	500.0



# 활동 과정 ↗

제주 특산물 가격 예측 AI 경진대회 (상위 3%)

## 8회차

- 데이터 분석 결과 최종 검증
- 데이콘에 결과 제출
- Score 확인 및 성능 개선 방안 모색

baseline\_submission (26).csv

day\_weight 열 추가 : 평일 2, 토요일 1, 일요일 0

680.5063448422

model = RandomForestRegressor(n\_estimators=400, min\_samples\_

일요일에 price가 있는 행 제거 edit

baseline\_submission (25).csv

real\_holiday 열 추가

724.4215919211

파라미터 변경

model = RandomForestRegressor(n\_estimators=400, min\_samples\_

edit

## 9회차

- 최종 리더보드 순위 확인
- 추가적인 대회 참여 여부 결정

Rank	Score	Submit
36 paradise	pa 871.67174	29

day\_weight 열 추가 : 평일 2, 토요일 1, 일요일 0

model = RandomForestRegressor(n\_estimators=400, min\_samples\_split = 2, min\_samples\_leaf = 1, max\_features = None)

일요일에 price가 있는 행 제거 edit



# 활동 과정

대구 교통사고 피해 예측 AI 경진대회 (상위 25%)

## 10회차

- 대구 교통사고 피해 예측 AI 경진대회 참여
- 데이터 개요 파악
- 데이터 시각화 및 전처리
- 데이콘에 결과 제출 및 피드백
- 향후 스터디 방향 논의

Rank

190

seomj0

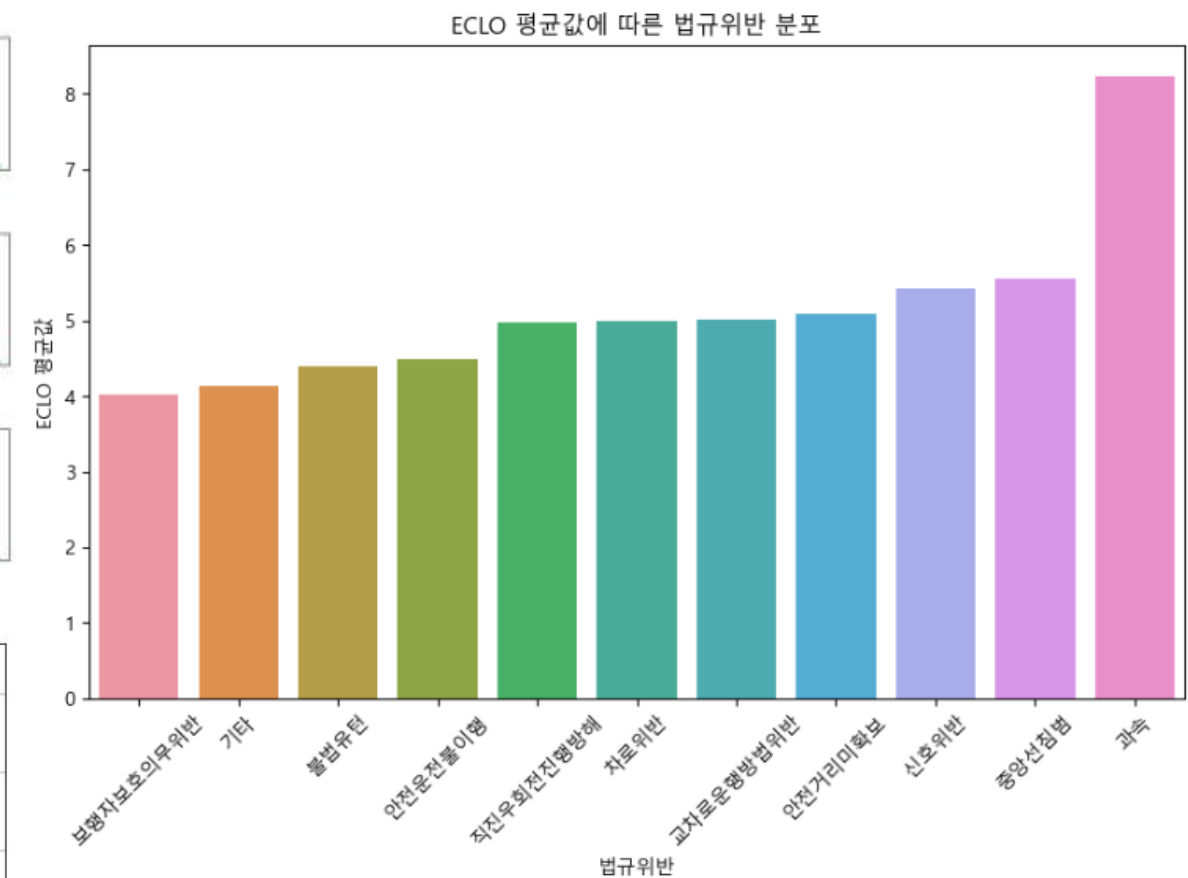
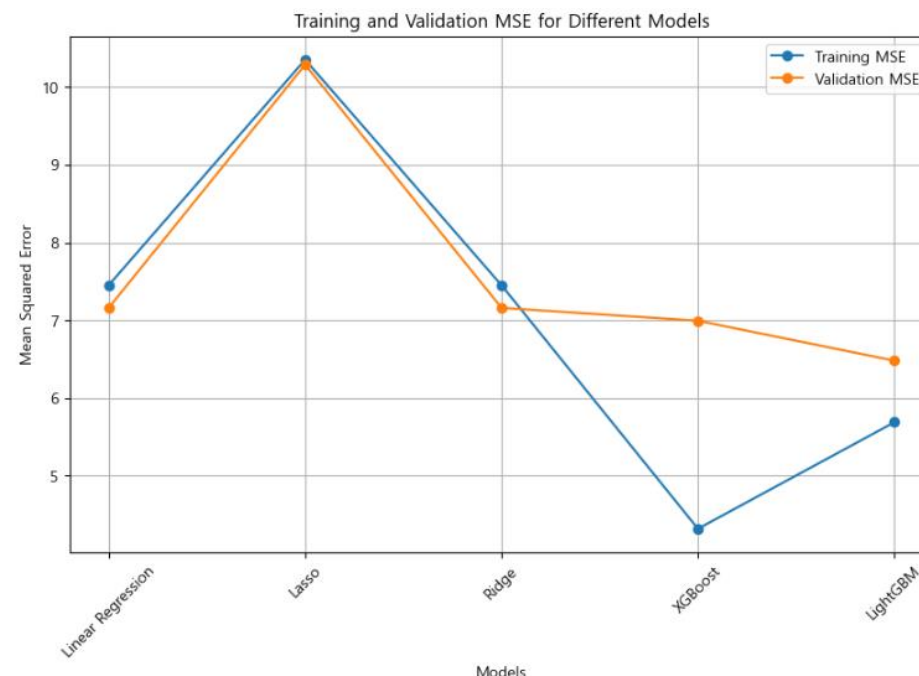
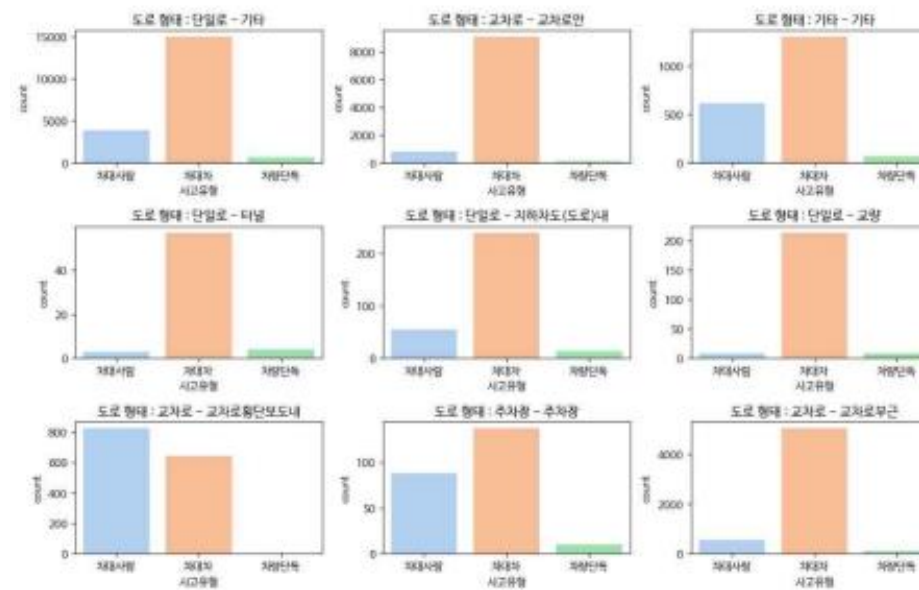
Score

se

0.42851

Submit

14



# 활동 과정

백엔드 세미나 및 실습 진행

## 11회차

- 백엔드 자체 세미나 및 피드백
- AI에서의 백엔드 활용 방안 토의



## 12회차

- 백엔드 개발 환경 구축 및 실습
- 최종 결과물 제작을 위한 역할 분담 및 논의

```
C:\workspaces\nodejs\server_basic>node server
Server is running...

localhost
localhost:3000
Hello node.js!!

try:
# 데이터베이스에 새로운 데이터 추가
with connection.cursor() as cursor:
    for index, row in fake_df.iterrows():
        insert_data_query = f"INSERT INTO {table_name} (id, saturation,
            f"VALUES ({row['id']}, {row['saturation']},
            f'{row['past_time']}', {row['seat']})"
        cursor.execute(insert_data_query)

# 데이터베이스에서 id를 기준으로 오름차순으로 정렬하여 조회
sorted_data_query = f"SELECT * FROM {table_name} ORDER BY id"
sorted_df = pd.read_sql(sorted_data_query, connection)
print(sorted_df)

# 커밋
connection.commit()

finally:
# 연결 닫기
connection.close()
```



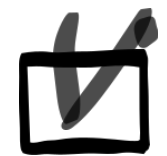
# 활동 과정 요약



## HD현대 AI Challenge

- 사용한 기법: OLS, 상관관계 분석
- 사용한 모델: Dense, H2O, TPOT, Google Cloud
- 결과: 117/330, 상위 35%

Rank		Score	Submit
117	seomj0	45.90871	24



## 대구 교통사고 피해 예측 AI 경진대회

- 사용한 기법: 자연어처리
- 사용한 모델: RF, Dense
- 결과: 190/880, 상위 21%

Rank		Score	Submit
190	seomj0	0.42851	14



## 제주 특산물 가격 예측 AI 경진대회

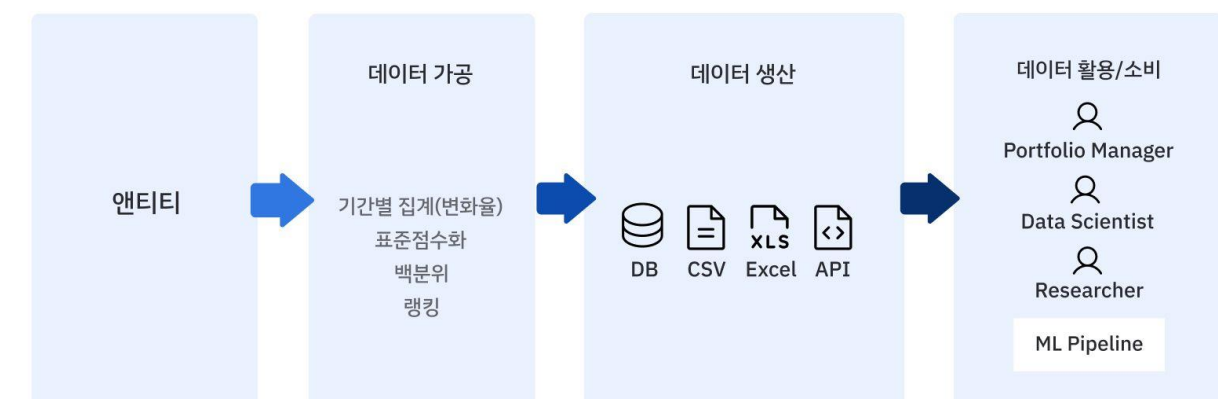
- 사용한 기법: 시계열 특성에 따른 데이터 시각화/전처리
- 사용한 모델: RF, Prophet, Autogluon
- 결과: 36/1093, 상위 3%

Rank		Score	Submit
36	paradise	871.67174	29



## 백엔드 세미나 및 실습 진행

- 사용한 프레임워크: Spring boot, Node.js
- 결과: 서버 구축 및 데이터 수집/가공 과정 이해





# 학습 성과

## 학습 목표

- 인공지능을 활용한 데이터 분석 및 예측 경험
- 예측 성능보다는 데이터 분석, 예측 기법에 대한 이론적 내용과 필요성 인식
- Pandas, numpy, scikitlearn 등의 분석 기법과 수학적 이론을 기반으로 폭넓은 분석과 다양한 인사이트 도출
- 백엔드의 이해를 통한 인공지능에 활용되는 데이터 수집 및 가공 방식의 이해

## 학습 성과

- HD현대 AI Challenge, 제주 특산물 예측 AI 경진대회, 대구 교통사고 피해 예측 AI 경진대회 참여
- 각 대회에서 상위 35%, 3%, 21%의 순위로 마무리
- 데이터의 특성(시계열, 수치/범주형 등)에 맞춘 데이터 분석 및 시각화 방법 숙지
- H2O, RF, Prophet, Autogluon, SVM 등 다양한 모델의 특징과 강점 이해
- 데이터 수집과 가공에 초점을 맞춘 백엔드 스터디





# 결론 및 느낀 점

## 결론 및 느낀 점

- 처음 사용해보는 모델을 이해하는 시간이 길어졌지만, 다양한 모델의 특성과 성능을 분석하는 과정에서 풍부한 경험을 쌓을 수 있었다.
- 각 모델의 파라미터의 이해와 데이터 전처리의 중요성을 깨닫게 되었다.
- 데이터와 관련된 백엔드 공부 및 서버 구축에 대한 실습 활동에서 새로운 경험을 얻을 수 있었다.
- 앞으로의 학습과 프로젝트에서도 이번 경험을 활용하여 최상의 결과물을 만들 수 있을 것이라는 자신감을 얻었다.

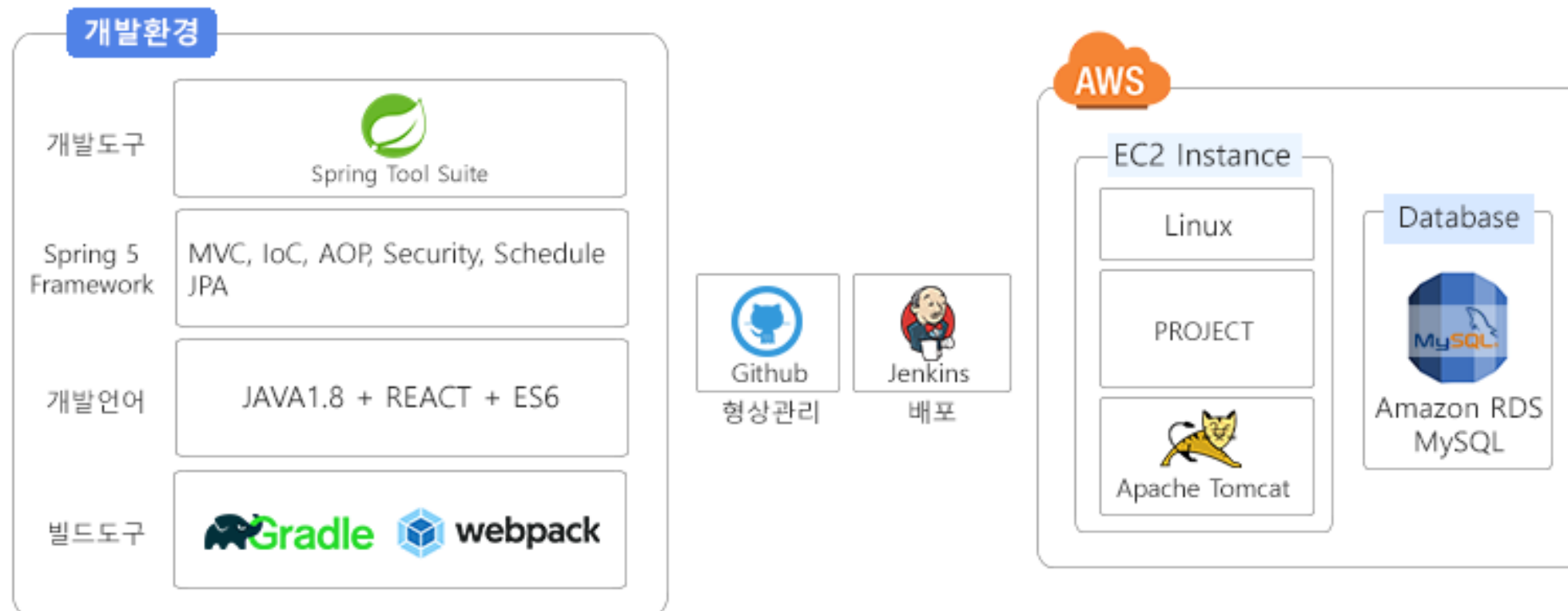




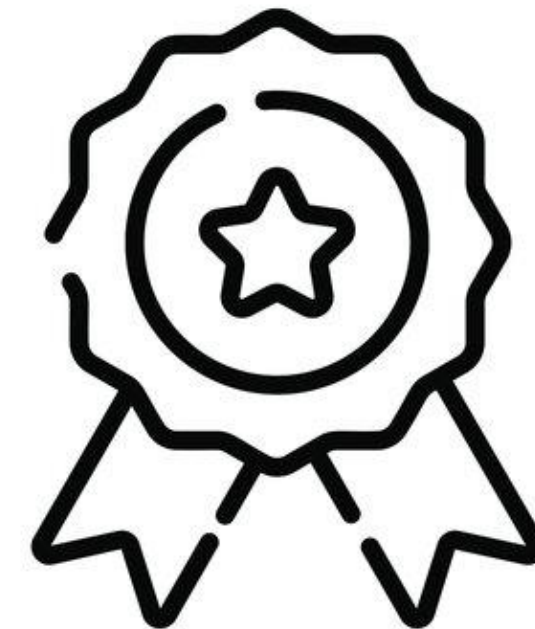
# 앞으로의 계획

## 앞으로의 계획

- 지금까지 공부한 AI 및 백엔드 지식을 활용한 프로젝트 진행
- 최종 목표는 관련 공모전 참가 및 수상



출처: [https://blog.naver.com/mering\\_k/221530087983](https://blog.naver.com/mering_k/221530087983)





# 감사합니다

70.0

컴퓨터공학과 21011591 서민정

컴퓨터공학과 19011502 최락원

컴퓨터공학과 19011558 서상민

