

『 최종 프로젝트 수행 계획서 』

제 출 일	2021.11.1	담당 강사	임동조
팀 명	삼삼사자	팀 인원	3명
기술스택	◆ 머신러닝 알고리즘(데이터 EDA, 머신러닝 모델 구축 및 평가) <ul style="list-style-type: none"> ➢ 지도학습 <ul style="list-style-type: none"> - 회귀(linear regression, polynomial regression), 의사결정트리, 앙상블(랜덤 포레스트, 그래디언트 부스팅(lightgbm, xgboost, CatBoost)) ➢ 비지도학습 <ul style="list-style-type: none"> - K-means, DBSCAN, PCA ➢ 시각화 라이브러리 <ul style="list-style-type: none"> -matplotlib, seaborn, plotly 		
프로젝트명	데이콘 - 가스공급량 수요예측 모델개발 대회에 참여		

■ 프로젝트 목표 및 개요

- ◆ 대회명 : 데이콘 가스공급량 수요예측 모델개발 대회
- ◆ 대회 링크 : <https://dacon.io/competitions/official/235830/overview/description>
- ◆ 주제 : 한국가스공사의 시간단위 가스 공급량 데이터와 기상 데이터 및 유가 데이터를 종합한 데이터셋을 구축하여 90일 한도 일간 공급량을 예측하는 인공지능 모델을 개발
- ◆ 주최 및 주관 : 한국가스공사
- ◆ 대회 개요: 한국 가스 공사가 보유한 다년간 시간 단위 공급량 데이터를 기반으로 미래 공급량을 예측하는 모델을 만든다.
- ◆ 외부 데이터 :
 - ◆ - 유가 데이터 - <https://www.data.go.kr/dataset/15004612/fileData.do>,
 - ◆ - 날씨 데이터 - <https://www.weather.go.kr/w/obs-climate/land/past-obs/obs-by-day.do?stn=184>
- ◆ 프로젝트 목표 :
 - 머신러닝 과정에서 배운 내용을 토대로 실전 분석 대회에 참여하여 기술의 응용력을 높임으로서 실전 역량을 키운다,
 - 가스 공급량을 예측하는 머신러닝 모델을 구축하고 다양한 모델의 비교를 통해 최적의 모델을 찾는다.

■ 역할 분담

성 명	분담 내용	역 할
박승규	<ul style="list-style-type: none"> - 데이터 (날씨, 유가, 가스 공급량) 병합 및 결측치 처리 - feature Engineering을 통한 변수 생성(통계량을 이용한 특성생성, 모델을 이용한 특성 생성), 모델 성능 개선 - 모델 평가(교차 검증, 그리드 리서치 함수 등을 활용) - plotly를 이용한 데이터 시각화, 발표 자료 준비 	팀장
이성준	<ul style="list-style-type: none"> -MSE, RMSE를 활용하고 NMAE평가 지표를 만든 후 모델 최종 평가 -교차 검증 및 파라미터 튜닝(max_depth, n_estimators, max_features), test_size를 크기를 달리해서 비교 - matplotlib, seaborn 등을 활용한 데이터 탐색 및 시각화 - ppt 자료 정리 	부팀장
노현곤	데이터 수집 <ul style="list-style-type: none"> - 데이터 탐색 및 수집 	팀원

■ 예상 결과물

- (1) 데이터 탐색 및 시각화 결과물
- (2) 최종 정리 결과물
- (3) 모델별 비교 평가 결과
- (4) 소스 코드 자료