텍스트이(가) 표시된 사진

자동 생성된 설명

**『 최종 프로젝트 수행 계획서 』**

멋쟁이사자처럼 K-Digital Training 『인공지능 통합과정』

|  |  |  |  |
| --- | --- | --- | --- |
| **제 출 일** | 211102 | **담당 강사** | 임동조 |
| **팀 명** | 으쌰으쌰 | **팀 인 원** | 김범중 김진연 |
| **기술스택** | 외부 데이터 크롤링, 랜덤 포레스트, 결정 트리 모델 생성 예정 | | |
| **프로젝트명** | 날씨에 따른 가스공급량 수요예측 모델 개발 | | |

**■ 프로젝트 목표 및 개요**

[목표]

- 실전 데이터를 활용한 머신 러닝 및 딥러닝 모델 구축 및 실전 분석 능력 향상

- 상위 5%에 들기

- 머신 러닝 및 딥러닝 원리 이해, 활용 및 기술 습득

[개요]

(1) 데이터 탐색 및 시각화 (태블로, matplotlib, seaborn, plotly)

(2) 외부 데이터 수집 및 데이터 병합

- 시간별 날씨 데이터 크롤링 (나주시 농업기상정보시스템)

(3) 머신 러닝 모델 구축 및 평가

- 데이터 전처리

- k-fold 교차 검증

- 회귀(선형회귀, Ridge, Lasso)

- knn, 의사결정트리, 앙상블(랜덤포레스트, lightgbm, xgboost, catboost 등)

- pycaret를 활용한 모델 비교 평가

[대회 정보]

일정: 데이터 확인 - 머신 모델 테스트 - 모델 선택 – 개선

* 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

■ **역할 분담**

|  |  |  |
| --- | --- | --- |
| **성 명** | **분담 내용** | **역 할** |
| 가채원 | - 프로젝트 전반적인 계획 및 팀 조율  - 데이터 탐색 및 시각화  - 중간 및 최종 결과 발표 자료 작성 (노션 정리)  - 자료 리서치  - 데이터 탐색 및 시각화(태블로 활용) | 팀장 |
| 김범중 | - 데이터 수집 탐색 및 전처리  내용 : 나주시 농업기상정보시스템의 기온 데이터  URL : https://weather.naju.go.kr/main/main.html  - 모델 구축 및 평가  평가 지표(MSE, RMSE, 결정계수, MAE, NMAE)를 활용한 평가  - 모델 개선 | 부팀장 |
| 김진연 | - 데이터 탐색 및 시각화  - 시각화 지원(태블로)  - 모델 구축 및 개선  - 회귀(선형회귀), 의사결정트리  - 앙상블(rf, xgboost, lightgbm, catboost 등)  - pycaret를 활용한 모델 비교 | 팀원 |
| 윤진훈 | - 데이터 탐색 및 시각화  - 모델 구축 및 개선  - 회귀(선형회귀), 의사결정트리  - 앙상블(rf, xgboost, lightgbm, catboost 등) | 팀원 |
|  |  |  |

■ **예상 결과물**

날짜에 따른 가스 공급량 예측에 대한 결과물

(1) 최종 프로젝트 소스 코드

(2) 최종 프로젝트 발표 자료

(3) 모델 비교 평가와 파라미터 변경 이력 등의 엑셀 정리 자료