**프로젝트 기획안**

| 팀 명 | 수요예측미식회 |
| --- | --- |
| 팀 원 | 이지예, 이예주, 김영준, 김남이 |
| 프로젝트 주제 | 기업 구내식당 식수 인원 예측 AI를 통한 인사이트 도출 |
| 프로젝트 수행 방향  (주요 기능 설명) | **배경**: DACON 경진대회 내 참신한 주제 및 AI 모델링에 적합한 주제를 선정 ([구내식당 식수 인원 예측 AI 경진대회 - DACON](https://dacon.io/competitions/official/235743/overview/description))  **목적:**   1. 구내식당 식수 인원 예측을 통한 효율적인 구내식당 운영 2. 비용절감을 통한 양질의 식사 제공 (직원 만족도 향상, 수요 반영) 3. 환경 문제에 기여하여 기업 브랜드 이미지 제고 (기업 PR 효과 )   **필요 작업:**   * 요일별 식수인원 비율 파악 :   (식사 가능 인원 수 대비 실제 구내 식당 이용 인원으로 계산)   * 메뉴와 식수인원 상관관계 파악:   (해당 과정에서 유의미한 상관관계 발견시, 인기 메뉴 추출 후 워드클라우드로 시각화)   * 반복되는 메뉴 제외 (ex. 쌀밥, 김치 등) * 외부 변수 고려(날씨, 공휴일, 연차, 외근, 재택근무 등)   **작업 상세 계획 :**   1. 데이터 랭글링 (KNN 알고리즘 활용 결측치 Imputing) 2. Feature Engineering   - 날짜 파생변수 생성1 (반기, 분기, 계절, 공휴일 전후일여부)  컬럼 : 날짜, 요일  할일1 날짜에서 계절 컬럼 생성  - 날짜 파생변수 생성2 (sin 및 cos 신호주기변환)  이건 할 지 안 할 지 고려 참고 -> ([시계열 예측 | TensorFlow Core](https://www.tensorflow.org/tutorials/structured_data/time_series?hl=ko))  스크롤 내려보면 sin, cos 주기신호변환 하는게 있음  유의미할지 안할지는 해보고 시각화 후 판단해야댈듯  - 메뉴 임베딩 (일반밥 0 덮밥류, 비빔밥류 1 인코딩)  *메뉴를 어떻게 인코딩 할 것인지?*  *- 밥 국 반찬1,2,3 디저트 김치 (0,1,2,3,4,5,6)*  *- 덮밥류(비빔밥류)등에 대한 7번 변수로 추가 ->중식계 비율 측정해서 상관관계*  - 식사 가용 인원 파생변수 생성  *계산식) 본사정원수 - (휴가자수 + 출장자수 + 재택근무자수)*  *중식계/가용인원 , 석식계/가용인원 비율을 측정해서 특정 메뉴에 대한 선호 파악 가능*   1. EDA 및 변수영향파악  * 요일 변수 (막대그래프): 월~금 실수요와 평균 * 연휴 변수 ‘1’로 측정된 값들과 요일 평균을 비교 * 2020 이전/이후 실수요 비교 * 특식과 중식계 간의 상관관계 * 신메뉴와 중식계 간의 상관관계 * 날씨 변수(강수량:max, 기온:mean) * 중식 시간대 11시~13시 (석식 17시~19시)) - 산점도  1. Feature Selection (필요시) 2. 모델링 3. 성능평가 및 모델선택 4. 비즈니스 인사이트 도출 (+ 클러스터링 분석) |
| 역할 분담 및 일정 | **역할 분담 :**  이지예: 데이터전처리, 시각화, 모델링, 팀장1  이예주: 데이터전처리, 시각화, 모델링, 팀장2, PPT제작  김영준: 데이터전처리, 시각화, 모델링, 팀장3, 발표  김남이: 데이터전처리, 시각화, 모델링, 팀장4, 발표  **일정 :**  9/8(수): 주제 선정 및 기획안 작성  9/9(목) ~ 9/10(금): 데이터 전처리  9/13(월): 시각화, 모델 학습 계획  9/14(화) ~ 9/15(수): 모델 학습 및 인사이트 도출  9/16(목): 발표자료 마무리 및 연습 |
| 프로젝트 수행 도구 | **<PYTHON>**  판다스, 넘파이, Scikit-learn(선형모델 및 KMeans), 머신러닝 패키지 (XGBoost, LightGBM, CatBoost 등), 텐서플로우(MLP), Matplotlib 및 Seaborn(시각화), statsmodels(시계열 분석)  **\*Regression (방법 추천)**  다항회귀분석, Lasso, Ridge, ElasticNet 등을 이용  딥러닝-선형회귀분석  완전탐색-그리드서치  랜덤포레스트계열  부스트계열  XGBoost, LGBM  Tensorflow |