데이터 전처리 결과서

# **AI 푸드닥터(푸닥) 데이터 전처리 결과서**

* **팀 정보**
  + 팀원: 서주혁, 대성원, 신동익, 윤정연
  + 깃허브: <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN07-FINAL-3Team>
* **데이터 전처리 개요**
  + 파일 형식: csv
  + 데이터 수집일: 2025년 3월
  + 데이터 규모: 14,751건
  + 전처리 도구: Python
* **데이터 전처리 과정**

**1.1. 컬럼 선정**

* 데이터 분석에 필요한 주요 컬럼 선별
  + 식품명, 데이터구분명, 식품기원명, 식품대분류명, 대표식품명, 식품중분류명, 식품소분류명
  + 영양성분: 영양성분함량기준량, 에너지(kcal), 탄수화물(g), 단백질(g), 지방(g), 당류(g), 식이섬유(g), 나트륨(mg)
  + 기타: 식품중량

**1.2. 결측치 및 중복 제거**

* 결측치 처리: 필수 컬럼(식품명, 영양성분 등)에 결측값이 있는 행 삭제
* 중복 데이터 처리: 동일한 식품명이 반복되는 경우 제거
* 불필요한 데이터 제거: 분석 목적과 관련이 낮은 식품 제거
  + 제외된 식품군: 음료 및 차류, 장류/양념류, 유제품류 및 빙과류, 곡류/서류 제품

**1.3. 대분류명에 따른 식단 항목 매핑**

| 식단 항목 | 대분류명 |
| --- | --- |
| 밥 | 밥류 |
| 메인 반찬 | 볶음류, 조림류, 구이류, 찜류, 만두류 |
| 서브 반찬 | 튀김류, 전·적 및 부침류, 빵 및 과자류 |
| 서브 반찬2 | 나물·숙채류, 생채·무침류 |
| 국 | 국 및 탕류, 찌개 및 전골류, 죽 및 스프류, 면 및 만두류 |
| 김치 | 김치류, 장아찌·절임류, 젓갈류 |
| 기타 | 위 항목에 포함되지 않는 식품 |

* + **데이터 전처리 결과**
    - 최종 데이터 건수: 552
    - 식품별 영양성분 데이터 정제 완료
    - 식단 카테고리 매핑 완료
  + **향후 사용 계획**
  + **영양 분석**
    - 각 식단 항목별 평균 영양성분 분석
    - 특정 영양소(나트륨, 당류 등) 기준 초과 여부 확인  
      * **추천 시스템 개발**
    - 사용자의 영양소 필요량에 따른 맞춤형 식단 추천
      * **데이터 시각화 및 리포트 생성**
    - 대분류별 영양 성분 분포 시각화
    - 주요 영양 성분 트렌드 분석