Data Science Group 1

End to End 7process

1. Object Setting

**Data set Motivation**

: **Food waste** has become a **critical issue** in achieving the Sustainable Development Goals (SDGs) in today’s world. This project aims to systematically **analyze food waste** through data, and from this analysis, **propose efficient operational strategies and waste reduction plans**.

**Furthermore**, it explores the potential to implement an incentive system **based on the amount of food waste generated**, enabling restaurants or operational units to receive rewards according to their waste reduction performance, thus providing a practical and scalable management approach.

+   
일단 코드에 사용된 클러스트링과 regressor을 설명해자면,

클러스터링 기법은 KMeans를 사용했고, 이 클러스터링 기법을 사용한 이유는 이 데이터 셋에서 라벨이 없는 데이터에서 데이터의 패턴이나 군집을 자동으로 식별하고 싶었기 떄문.

회귀모델은 LGBMRegressor를 사용하였고, 특징은 Gradient Boosting 기반의 트리 모델이라는 것이고, 위 데이터 셋에서 예측할 목표값이 연속적인 수치(food\_waste\_kg)라서 사용했습니다

A graph of food waste distribution

AI-generated content may be incorrect.

위 이미지가 시사하는 바를 적자면, 다양한 유형의 환경이 있다는 것.

세개의 클러스터가 서로 다른 분포를 가짐.

🡪 이는 서로 다른 조리 환경, 인력 수준, 식사 수, 기후 조건 등이 음식물 쓰레기 발생에 영향을 준다.

* **Cluster 0**: 예측값이 45~55kg에 매우 집중 → **운영이 일관적이고 통제된 환경**
* **Cluster 1**: 30~70kg 분포, 조금 더 퍼져 있음 → **변동성 있는 일반적 급식 환경**
* **Cluster 2**: 긴 꼬리, 최대 120kg까지 → **특이하거나 대량 음식 공급 환경** (ex. 대형 행사, 미숙한 운영)
* Cluster 0은 낭비가 적고 예측이 안정적 → **효율적으로 운영되는 시설**일 가능성
* Cluster 2는 일부 극단적인 waste 예측 → **낭비가 심하거나 관리 미흡한 환경**이 섞여 있을 수 있음

정책적으로는 Cluster 2 유형의 식당을 타겟으로:

* **낭비 감축 캠페인**, **스마트 조리 시스템 도입**, **인력 교육 강화** 등 개입이 필요

A graph of a distribution of food waste

AI-generated content may be incorrect.

|  | **사회적 해석** |
| --- | --- |
| 40~45kg 대에 예측이 집중 | 낭비 기준선, 이 수준 이하가 “정상” 또는 “효율적”일 수 있음 |
| 긴 오른쪽 꼬리 | 소수의 급식소에서 **과도한 낭비 발생**, 정책 개입 대상 |
| 전체 분포는 비대칭 | **평균보다 중간값 기반 정책이 더 합리적** |
| 30~60kg에 대부분 존재 | **효율화 캠페인**, **정량 제공 교육**을 집중할 최적 범위 |

A graph of blue rectangular bars

AI-generated content may be incorrect.

| **클러스터** | **샘플 수** | **해석** | **제안** |
| --- | --- | --- | --- |
| 0 | 가장 많음 | 일반적이고 안정된 운영 형태일 가능성 | 모범 사례 전파 |
| 1 | 중간 수준 | 균형 잡힌 환경일 수 있음 | 평균 개선 전략 적용 |
| 2 | 가장 적음 | 특이 환경일 가능성, 낭비 리스크 있음 | **집중 개입 필요** |