

데이터 처리 프로세스 설계

활동 초기의 가장 주요한 활동은 다섯 가지 맛별 재료 선정 및 데이터 추출이었다. 다섯 가지 맛에 대한 알맞은 재료 선택은 프로젝트의 성공적인 진행을 위한 핵심적인 요소였다. 데이터 추출 후 전처리 과정은 최종적인 인공지능 모델의 정확도에 큰 영향을 미칠 것이다. 팀원들의 노력과 협력을 바탕으로 데이터 추출 및 처리 방식을 최적화하는 작업을 진행해야 한다.

전처리 과정에서는 먼저 NIR 분광기로부터 얻은 원시 데이터의 불필요한 column 제거와 필터링 작업을 진행해야 한다. 이후, 데이터의 정규화 및 표준화 과정을 거쳐 모델 학습에 적합한 형태로 변환해야 한다. 이렇게 전처리된 데이터는 MLP 모델에 입력되어, 다섯 가지 맛별 특성을 학습하게 할 수 있다.

MLP 모델을 사용하려 했지만, 6주 차 활동을 하며 데이터의 정규화 및 표준화 과정을 진행하며 모델을 전면 수정하였다. 데이터들의 병합 및 리쉐입 과정을 통해 CNN모델에 적용할 수 있도록 수정해야한다. NIR 분광기에서 추출된 CSV 데이터 파일은 모델링에서 CNN 모델에서 전처리 과정과 reshape과정을 거쳐 적용되었다.

우선, 5주차의 선행 연구에서 얻은 모델 관련 지식을 바탕으로 데이터 처리의 전반적인 흐름을 설계하였다. 데이터의 측정부터 전처리, labeling, 데이터 reshape, 그리고 모델 학습까지의 단계별 처리 방식을 검토했다. 특히, 선행 연구에서 언급된 CNN 작동 방식을 참고하여, 다섯 가지 맛별 데이터의 특징 추출에 CNN을 활용하기로 하였다.

