

데이터 수집 및 전처리

본격적인 프로젝트 시작에 앞서 분석의 유효성을 시각적으로 확인하고 모델을 구성하기 위해 샘플 데이터를 수집하였다.

3-1 개요

NIR 분광기를 통해 데이터를 수집하였다. 데이터를 정형화하기 위해 가루로 분쇄하여 사용하였다.

각 재료별 학습용 데이터로 1000개씩 수집하였다.

아래는 데이터 수집 장면이다.



3-2 파일 설명

white sugar.csv	백설탕
brown sugar.csv	황설탕
black sugar.csv	흑설탕
cube sugar.csv	각설탕
sugar powder.csv	슈가 파우더
iron.csv	철
paper cup.csv	종이컵
plastic.csv	플라스틱

stone.csv	돌
Himalayas salt.csv	히말라야 소금
cheon il yeom.csv	천일염
salt.csv	소금
seasoned salt.csv	맛소금
hub salt.csv	허브솔트
citric acid.csv	구연산
lemon powder.csv	레몬파우더
malic acid.csv	사과산
sour candy.csv	샤워캔디
stannic acid.csv	주석산
bitrex.csv	비트렉스
coffee.csv	커피가루
deodeok.csv	더덕
enchovy.csv	멸치가루
EP.csv	타이레놀
curry.csv	카레가루
dasima.csv	다시마가루
garlic.csv	마늘가루
mussel.csv	홍합가루
pyogo mushroom.csv	표고버섯가루

3-3. 전처리

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.

아래는 히말라야 소금의 전처리 후 모습이다

	A	B	C	D	E	F	G	H	I
1	704.38	704.87	705.35	705.83	706.32	706.81	707.29	707.78	708.27
2	28	26	28	27	27	28	27	28	28
3	28	27	27	27	29	27	28	29	28
4	28	28	28	29	29	29	28	28	28
5	30	28	30	30	30	30	30	29	32
6	28	29	31	27	30	30	29	31	31

