

# 용해탱크 제조 데이터를 이용한 품질 이상 탐지 및 진단

허준봉  
이성호  
야오와말

# 용해공정이란

## 용해공정이란?

- 분말 원재료를 액상 원재료에 녹이는 공정으로, 식품, 화학, 석유화학 등 다양한 분야에서 적용된다.
- 분말 유크림, 기능성 조제 분말 등을 생산하는 식품제조업의 용해공정으로, SD/MSD 건조생산라인의 원료 전처리 작업의 첫 번째 단계에 속한다.
- 살균, 분무건조 등 후공정에서의 공정품질을 보장하기 위해서는 다양한 분말 및 액상 원료가 균일한 혼합물이 되는 것이 중요하다.



용해탱크

# 1. 소개

- 배경

- 용해공정의 품질이 후공정 및 완제품의 품질에 미치는 영향이 크므로 품질을 잘 예측하는 것이 중요함.
- 동시간에 다수개의 원료를 투입하고 사람이 일일이 품질이 양품인지 아닌지 확인할 수 없음.  
따라서 비용과 시간을 줄이기 위해 품질예측 모델이 필요함.

- 목표

- 용해 품질 예측 모델링 만들기
- 용해 품질이 각 연속형 변수들이 어느 범위를 가질 때 품질이 정상(OK)가 나오는지 분석하기

## 2. 데이터

- 데이터셋 출처: 인공지능 중소벤처 제조 플랫폼
- 사용할 데이터: 독립변수( MELT\_TEMP, MOTORSPEED, MELT\_WEIGHT, INSP)와 종속변수( TAG)로 사용함

변수명	설명	데이터 타입
STD_DT	날짜, 시간 (YYYY-MM-DD HH:MM:SS)	Object
NUM	인덱스	int64
MELT_TEMP	용해 온도	int64
MOTORSPEED	용해 교반속도	Int64
MELT_WEIGHT	용해탱크 내용량(중량)	Int64
INSP	생산품의 수분함유량(%)	float64
TAG	불량여부	Object

## 2. 데이터 전처리

- 데이터에 결측치가 있는지 확인하기
- 데이터가 모두 양의 값을 가져야되는데 음수를 가진 데이터가 있으면 그 데이터의 행을 제거하기
- 데이터에 모델을 만들 때 필요없는 행이 있으면 제거하기
- 데이터가 불균형이어서 Smote 기법으로 데이터 수를 늘려주기
- 종속변수가 문자형이어서 범주형 변수로 바꾸어주기
- Train data, Validation data, Test data를 7대 2대 1로 나누기
- 변수들마다 단위가 다르므로 표준화 작업 진행하기

## 2. 데이터 탐색(EDA) 및 시각화

- 각 데이터에 대한 기술통계량을 확인하기
- 연속형 데이터 상관계수를 구하고 그것을 히트맵(heatmap) 그리기
- X축을 TAG(불량여부)로 하고 Y축을 연속형변수들로 정해서 Boxplot을 그림
- 변수 STD\_DT(날짜, 시간)는 시간을 나타내는 그래프이므로 X축을 STD\_DT로 놓고 Y축을 연속형변수들로 놓아서 꺾은선 그래프 그려보기
- 각 연속형 변수들이 시간에 따라 추이 분석하기
- 용해 품질이 각 연속형 변수들이 어느 범위를 가질 때 품질이 정상(OK)가 나오는지 분석하기

# 3. 분석 계획

## 1. 분석 기법

- KNN, Logistic Regression, Decision Tree, SVM, Naïve Bayes
- 앙상블 모형( Voting Classifier, Random Forest 등등 )
- LSTM 기법을 공부해서 할수 있으면 이상치 탐지해보기(미정)

## 2. 모델 평가 방법

- ROC Curve와 정오분류표로 가장 좋은 모델 선정 후에 test data에 평가해보기

## 3. 기대 효과

- 품질 분류기를 만들어서 비용과 시간을 절감할 수 있다.

## 4. 역할 분담

이름	역할
허준봉	데이터 전처리하기
	의사결정나무모형과 KNN 모형 만들고 하이퍼파라미터 튜닝하기
	PPT 만들기
이성호	데이터 시각화하기
	SVM과 베イズ분류기 모형 만들고 하이퍼파라미터 튜닝하기
	앙상블 모형 만들기
야오와말	데이터 탐색 및 시각화하기
	로지스틱 모형 만들고 시계열 Arima 모형 만들기
	품질이 정상일 때 각 변수들이 갖는 범위 분석하기