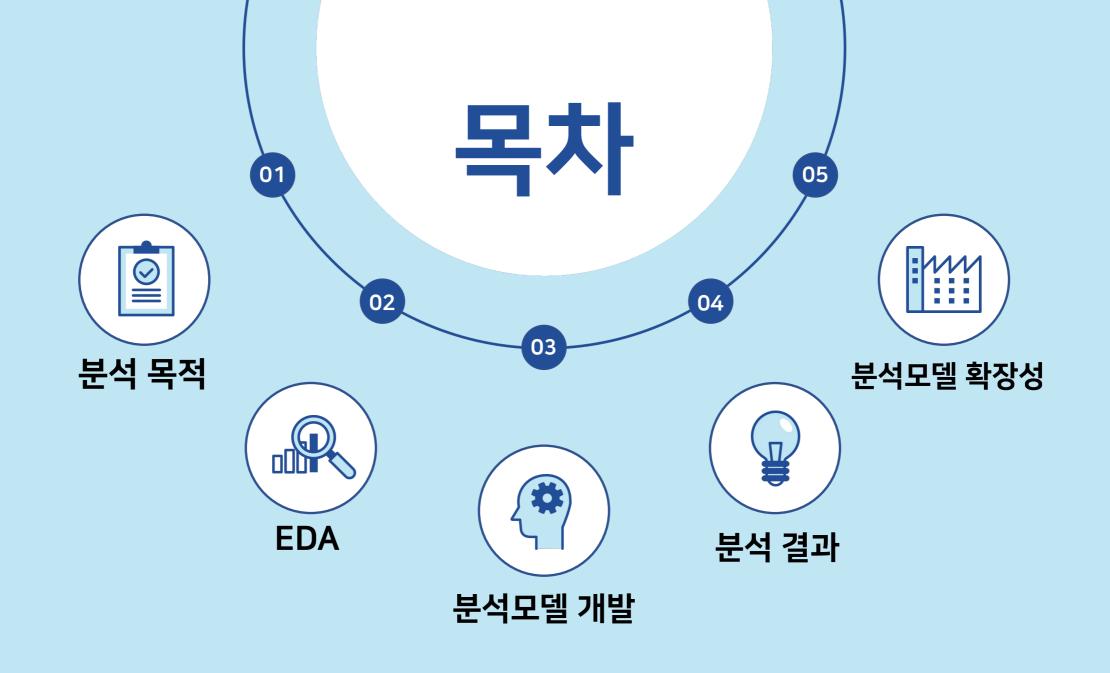


원료의 특성을 고려한 설명가능한 용해 공정 불량 예측 최적화 모델



01

분석 목적

- Target
- Problem
- Goal

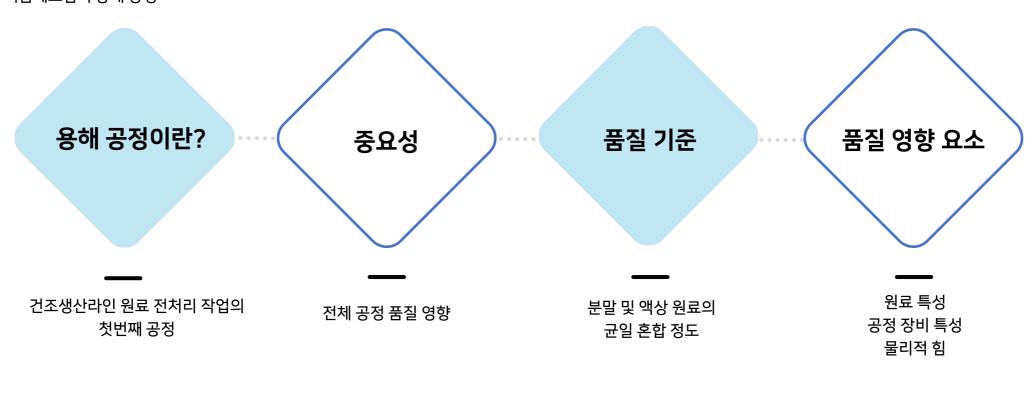


01 분석 목적



Target

식품제조업의 용해 공정



Problem투입 원료의 다양성 인력 부족

 \rightarrow

Goal

투입 원료의 특성 반영한 수율 예측 특성들이 품질 지표에 미치는 영향력 수치화



EDA

- EDA
- Data Preprocessing







• EDA

경진대회용 용해탱크 데이터셋.csv

STD_DT & NUM

MELT_WEIGHT

MELT_TEMP

MOTORSPEED

INSP

- 월 별 불량률 차이 ← 외부 환경 변동
- 초 별 불량률 차이 ← 공정 장비의 특성
- 일차 별 불량률 차이 ← 원료의 특성 변동

- MELT_WEIGHT > 2000 → 이상치
- 시간에 따른 증감 변동 주기 ← 원료의 특성
 - 주기 별 불량률 차이 ← 원료의 특성 변동

- 불량 존재0 구간 $308 \le MELT_TEMP \le 568$
- 불량 존재X 구간 676 ≤ MELT_TEMP ≤ 832
- 불량 존재이 구간 MOTORSPEED=0 54 ≤ MOTORSPEED ≤ 271
- 불량 존재X 구간 1669 ≤ MOTORSPEED ≤ 1804
- 일자 별 동일한 INSP의 불량률 차이
 - ← 원료의 특성 변동





• Data Preprocessing

One Hot Encoding

Clustering

Feature Generation Feature Selection

- INSP → 범주형 특성

- MOTORSPEED & MELT_TEMP → 혼합 분포 모델 군집 (←공정 장비의 특성) → CLASS_MOTORSPEED Input(CLASS_MOTORSPEED, SECOND, NUM)

- Output (TAG)

→ KNN 학습 → predict_proba() = PROBA_OK - 물리적 힘 특성 (MOTORSPEED, MELT_TEMP)

- 공정 장비 특성 (CLASS_MOTORSPEED)

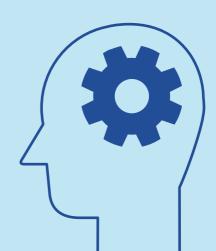
용해공정 품질 특성 (INSP)

- 공정 장비 및 원료의 특성 (PROBA_OK)

- 최종 완제품 품질 특성 (TAG) 03

분석모델 개발

- AutoML
- Bayesian Optimization
- XAI
- Data Split
- Model Architecture



03 분석모델 개발



AutoML

Feature → Pycaret AutoML → Model selection

best_models = compare_models(sort="Accuracy", n_select=3)

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
xgboost	Extreme Gradient Boosting	0.8413	0.9018	0.9452	0.8657	0.9037	0.4580	0.4727	11.8120
lightgbm	Light Gradient Boosting Machine	0.8389	0.8986	0.9554	0.8567	0.9033	0.4292	0.4533	0.5140
rf	Random Forest Classifier	0.8269	0.8856	0.9097	0.8755	0.8922	0.4531	0.4554	7.3580
gbc	Gradient Boosting Classifier	0.8208	0.8725	0.9638	0.8344	0.8944	0.3212	0.3626	10.8710
et	Extra Trees Classifier	0.8138	0.8590	0.8959	0.8714	0.8835	0.4209	0.4220	9.2690

→ XGBoost Model

Bayesian Optimization

Feature → XGBoost → Bayesian Optimization → Hyper-parameter optimization

```
{'target': 0.988855962643678, 'params': {'colsample': 0.885030598969056, 'gamma': 16.63782564380123, '11_reg': 566.0874092148872, '12_reg': 273.793769867424, '1r': 0.05740236271365283, 'max_depth': 9.736006499859627, 'min_child_w': 5.986727556541564, 'n_trees': 746.53513211 02312, 'subsample': 0.8861015002631764}}
```

XAI

- XGBoost → get_booster().get_score() → Feature Weight → Feature Importance
- SHAP → Feature Shapley Value

03 분석모델 개발

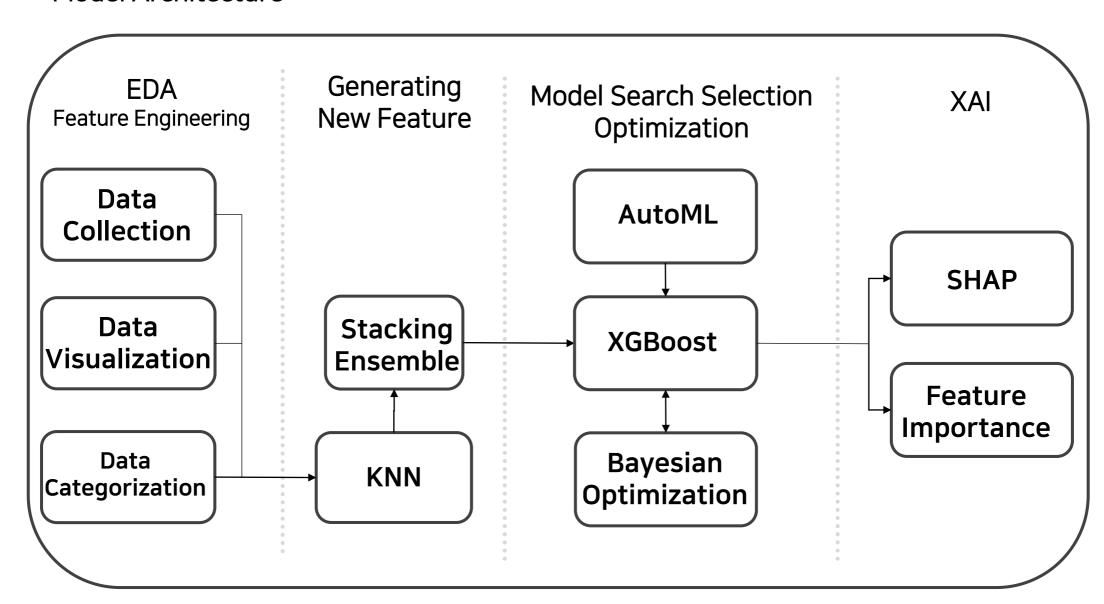


Data Split

KNN_Train_Data 40% **Train KNN** KNN_Validation_Data 10% **Optimize KNN** 32% **Train XGBoost** XGBoost_train_Data XGBoost_Validation_Data 8% **Optimize XGBoost** Test_Data 10% **Architecture Performance**



Model Architecture



04 분석 결과

- Architecture Performance
- Feature Importance
- SHAP



04 분석결과



Architecture Performance

XGBoost Train Data

XGBoost Validation Data

Test Data

<pre>from sklearn.metrics import classification_report y_pred_train = model.predict(X_train) print(classification_report(y_train,y_pred_train))</pre>						<pre>y_pred_val= model.predict(X_val) print(classification_report(y_val,y_pred_val))</pre>					<pre>y_pred_test= model.predict(test_x) print(classification_report(test_y,y_pred_test))</pre>				
princerassii	precision	recall	f1-score	support	р	recision	recall	f1-score	support		precision	recall	f1-score	support	
0	0.99	0.99 1.00	0.99 1.00	56662 210602	0 1	0.98 0.99	0.98 0.99	0.98 0.99	14165 52651	0 1	0.98 0.99	0.98 0.99	0.98 0.99	17707 65813	
accuracy macro avg weighted avg	0.99 1.00	1.00	1.00 1.00 1.00	267264 267264 267264	accuracy macro avg weighted avg	0.99 0.99	0.99 0.99	0.99 0.99 0.99	66816 66816 66816	accuracy macro avg weighted avg		0.99 0.99	0.99 0.99 0.99	83520 83520 83520	

Performance Comparison

KNN-XGBoost model

precision: 0.9948

recall: 0.9944

f1-score: 0.9946

accuracy: 0.9915

KNN-DNN model

precision: 0.9677

recall: 0.9902

f1-score: 0.9788

accuracy: 0.9662

Feature → AutoKeras → DNN Selection

LSTM model

precision: 0.9961

recall: 0.8063

f1-score: 0.8912

accuracy: 0.8055

GuideBook

04 분석결과

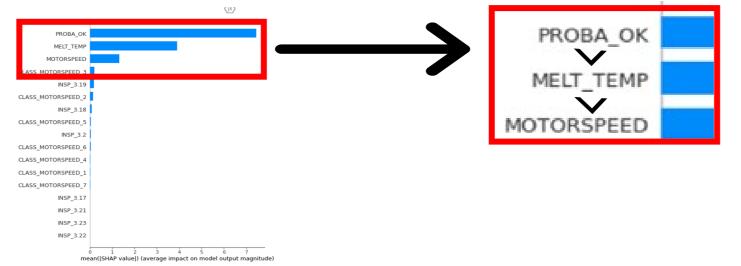


• Feature Importance



SHAP

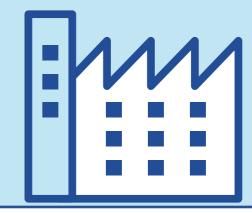
최종 품질에 대한 영향력 + 선형 관계성



05

분석모델 확장성

- 용해 공정에서의 발전 가능성
- 타 공정으로의 확장 가능성



05 분석모델 확장성



• 용해 공정에서의 발전 가능성



원료 조합 Labeling

- → 원료 조합별 PROBA_OK
- → 원료 교체 시, 초기값 설정
- → 초기 수율 예측 정확도 ↑

Feature Importance & Shapley Value

- → 용해온도 & 모터속도 기준 실험 계획
- → 용해온도 & 모터속도 최적 구간 도출

→ 품질 개선

장비 설정 값 및 불량에 따른 비용 최소화

→ 공정 최적화

05 분석모델 확장성



• 타 공정으로의 확장 가능성

식품 제조업의 가열 살균 공정

용해된 내용물 → 살균된 내용물

용해 내용물의 특성 + 장비 설정 값 →살균된 내용물의 품질 예측 철강 제조업의 고로/전기로 공정

- 고로 공정 철광석 + 석탄 → 용선(쇳물) - 전기로 공정 철 스크랩 → 용선(쇳물)

원재료의 특성 + 장비 설정 값 →용선의 품질 예측 바이오의약품 제조업의 배양 공정

균주/세포주 + 레진/미디어 → 균주/세포주 증식

균주/세포주 및 레진/미디어의 특성 + 장비 설정 값 →증식 결과 품질(양품/불량) 예측

Thank You 감사합니다

