

LevelUP

# 제조 데이터 분석

핀성형기의 가동 상태 탐지



하연주  
김유림  
김희주

# CONTENTS

1

분석 목적

2

데이터 설명

3

모델링

4

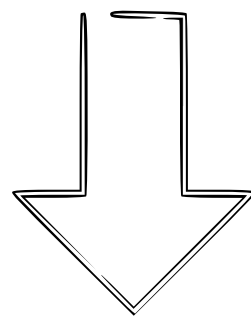
이슈 사항

5

결과

## 분석 목적

핀성형기 가동정보 데이터를 바탕으로 설치된  
센서 데이터를 확인하여 핀성형기  
START/STOP의 구분점을 찾는 것을 목표



각 데이터가 'START' 또는 'STOP' 중 하나로 분류되도록 모델을 학습



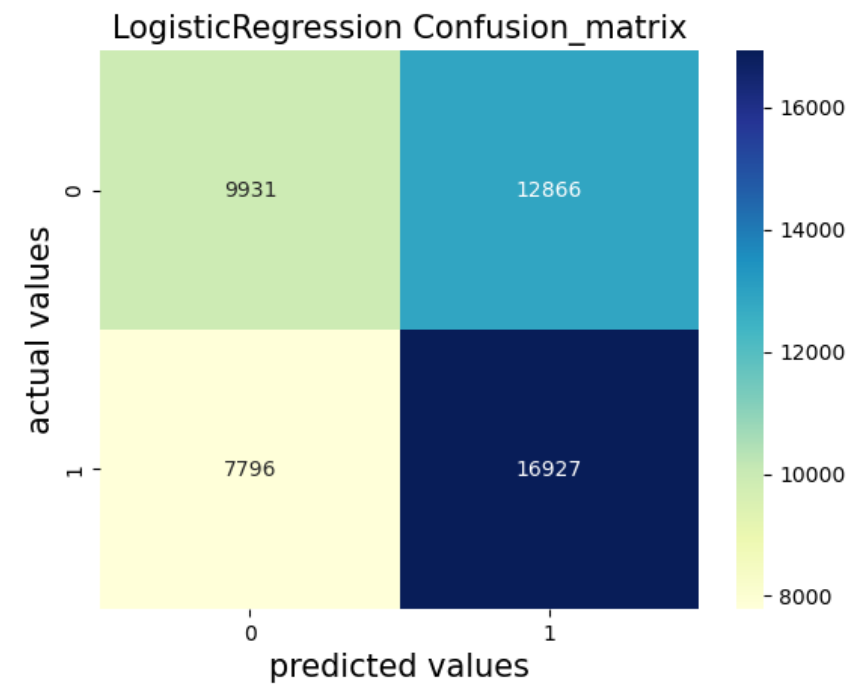
## 데이터 설명

	MAKING1_START	MAKING1_STOP	MAKING1_CNT
0	1	0	19978.679427
1	1	0	19978.679427
2	1	0	19978.679427
3	1	0	19978.679427
4	1	0	19978.679427
...	...	...	...
158395	1	0	37264.135652
158396	1	0	37264.135652
158397	1	0	37264.685514
158398	1	0	37264.685514
158399	1	0	37264.685514

158400 rows × 17 columns

- **MAKING1\_START: 가동** (단위 : ON/OFF, ON=1, OFF=0) (Type: int64)
- **MAKING1\_STOP: 정지** (단위 : ON/OFF, ON=1, OFF=0) (Type: int64)
- MAKING1\_CNT: 생산수량 (단위: EA) (Type: float64)
- MAKING1\_VOLT\_N상: N상 전압 (단위: V) (Type: float64)
- MAKING1\_VOLT\_AVG: 평균 전압 (단위 : V) (Type: float64)
- MAKING1\_CURR\_N상: N상 전류 (단위 : A) (Type: float64)
- MAKING1\_CURR\_AVG: 평균 전류 (단위 : A) (Type: float64)
- MAKING1\_KW: 유효전력 (단위 : KW) (Type: float64)
- MAKING1\_KVAR: 무효전력 (단위 : KVAR) (Type: float64)
- MAKING1\_HZ: 주파수 (단위 : HZ) (Type: float64)
- MAKING1\_PWR\_FACTOR: 역률 (단위 : %) (Type: float64)
- MAKING1\_KWH: 유효전력량 (단위 : KWH) (Type: float64)
- MAKING1\_KVARH: 무효전력량 (단위 : KVARH) (Type: float64)

# 모델링 - START



로지스틱 회귀

정확도 : 56.5%

정밀도 : 56.8%

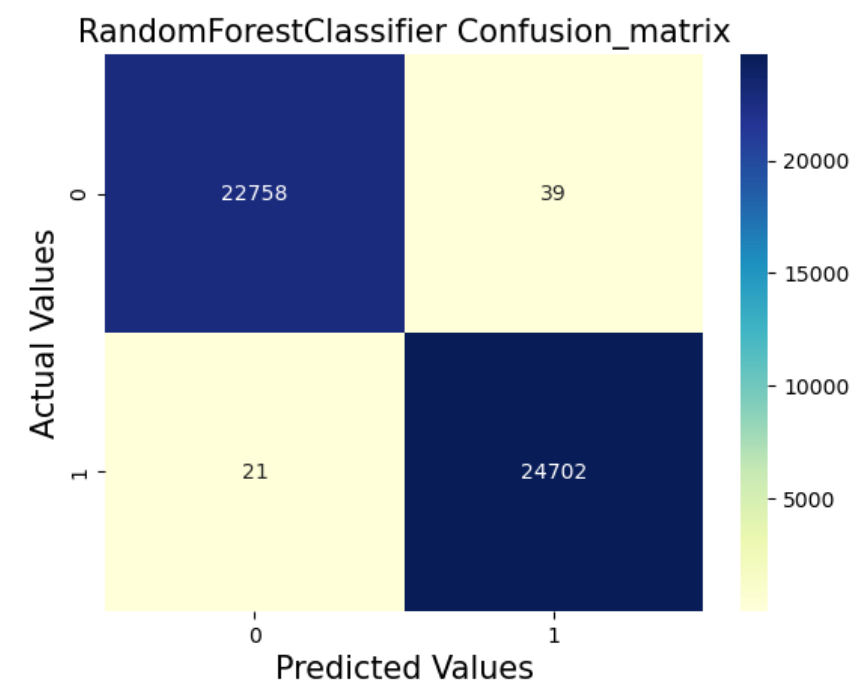
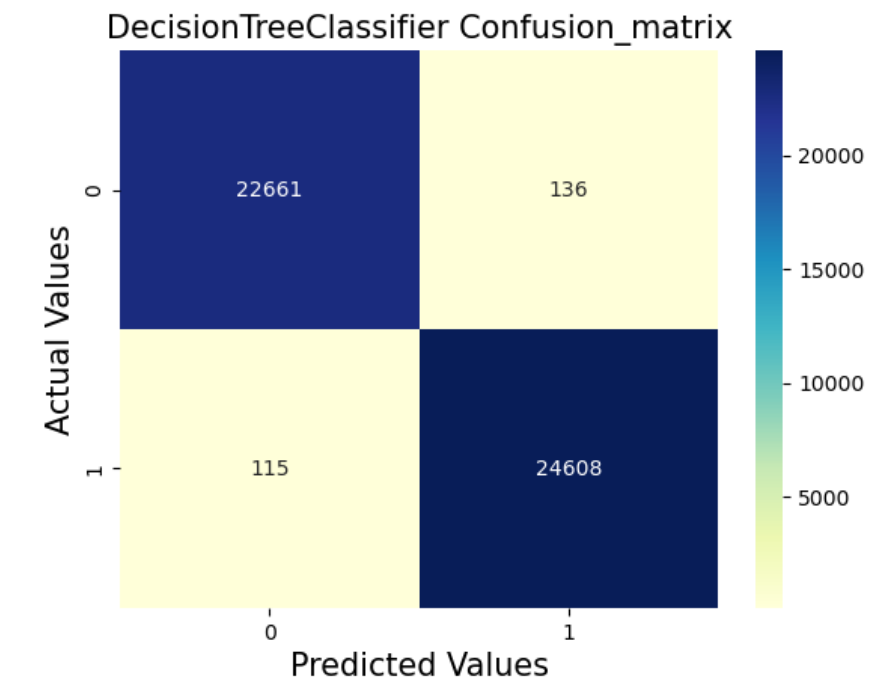
재현율 : 68.5%

의사결정나무 분류

정확도 : 99.5%

정밀도 : 99.5%

재현율 : 99.5%



랜덤포레스트 분류

정확도 : 99.9%

정밀도 : 99.8%

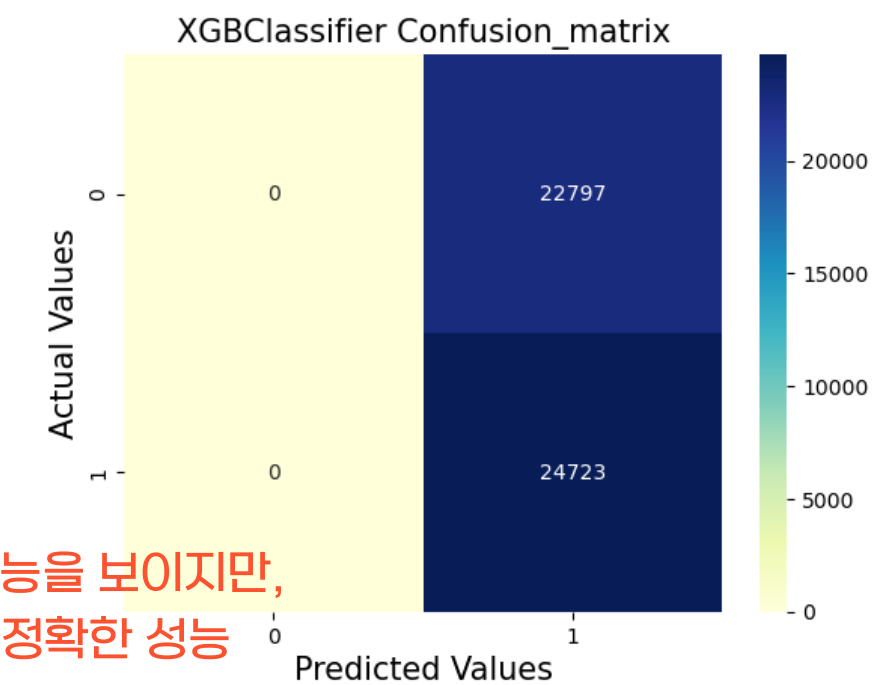
재현율 : 99.9%

XGBoost 분류

정확도 : 52.0%

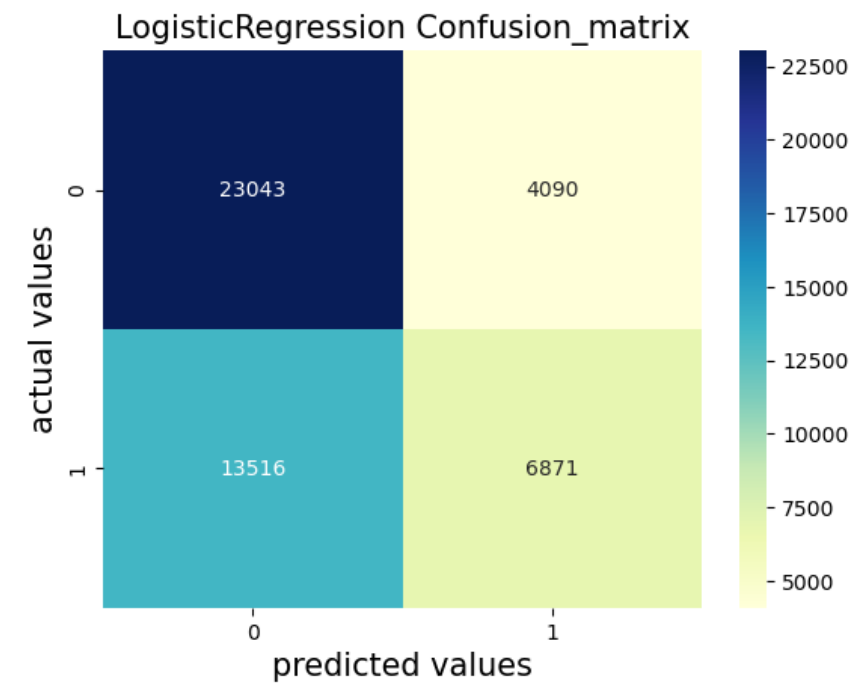
정밀도 : 52.0%

재현율 : 1.0%



Positive 클래스에 대해서는 매우 좋은 성능을 보이지만,  
Negative 클래스에 대해서는 매우 부정확한 성능

# 모델링 - STOP



로지스틱 회귀

정확도 : 63.0%

정밀도 : 62.7%

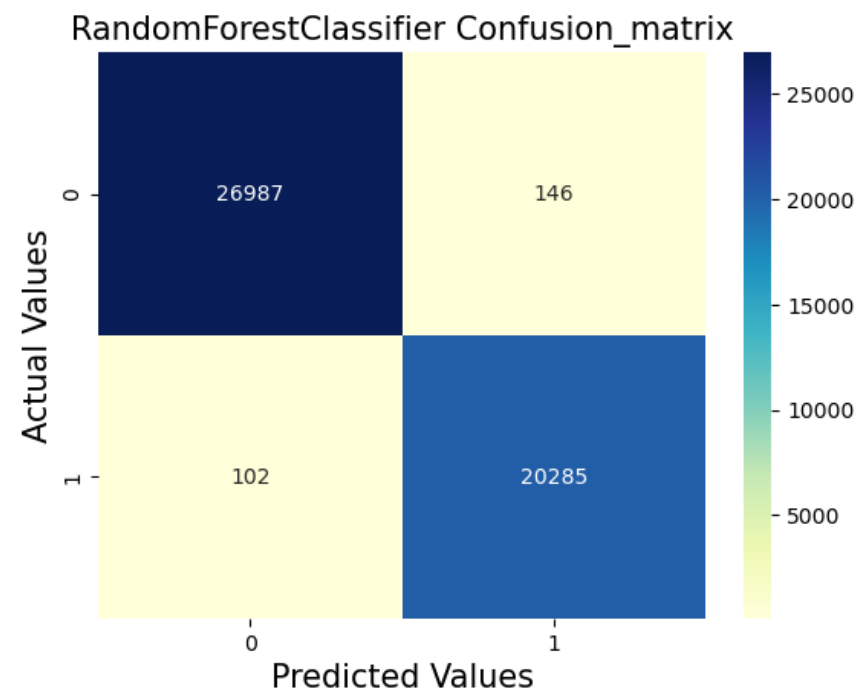
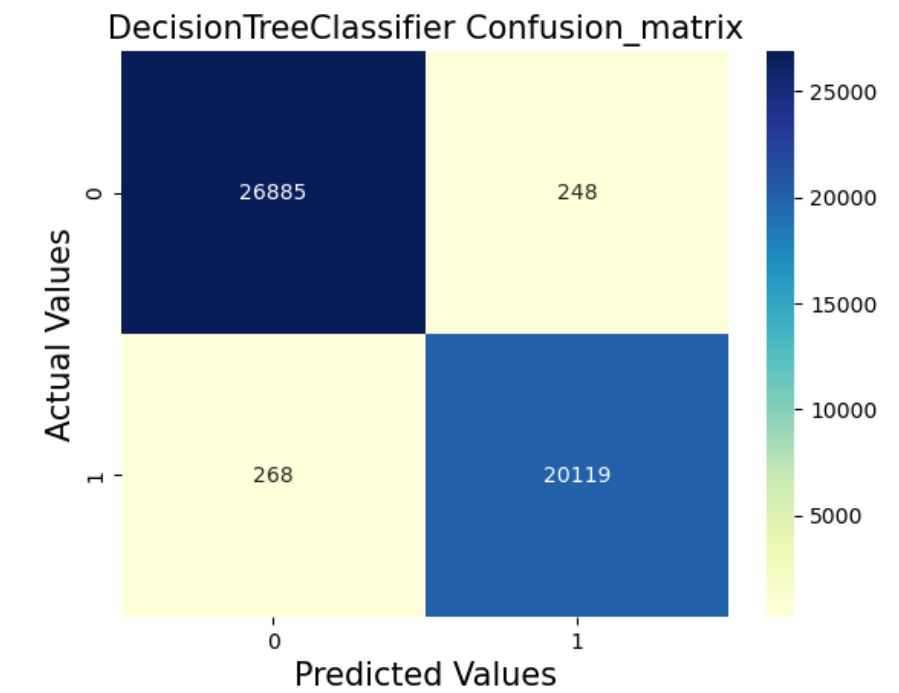
재현율 : 33.7%

의사결정나무 분류

정확도 : 98.9%

정밀도 : 98.8%

재현율 : 98.7%



랜덤포레스트 분류

정확도 : 99.5%

정밀도 : 99.3%

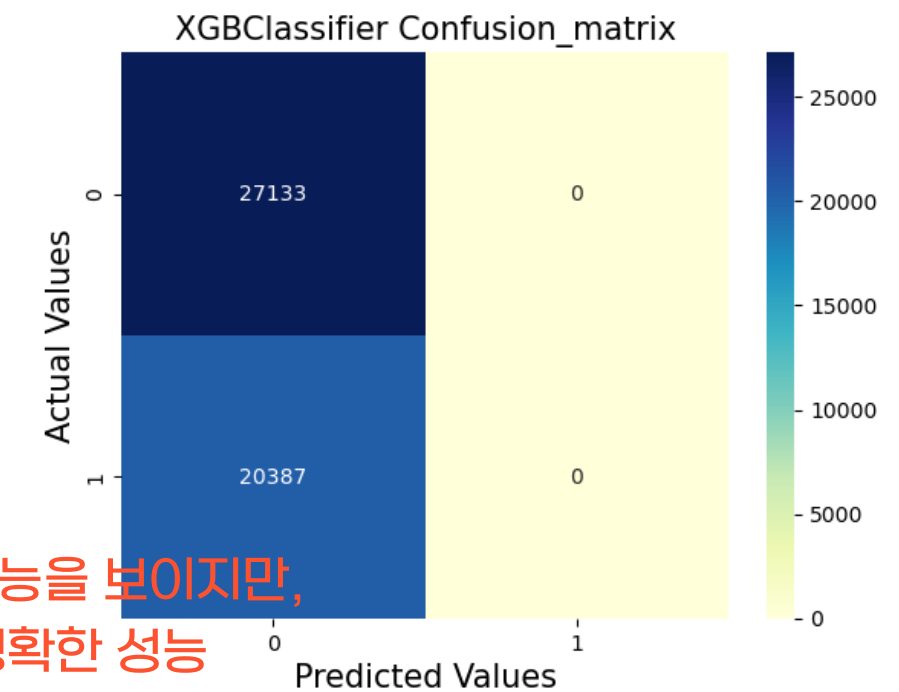
재현율 : 99.5%

XGBoost 분류

정확도 : 57.1%

정밀도 : 00.0%

재현율 : 00.0%



Negative 클래스에 대해서는 매우 좋은 성능을 보이지만,  
Positive 클래스에 대해서는 매우 부정확한 성능

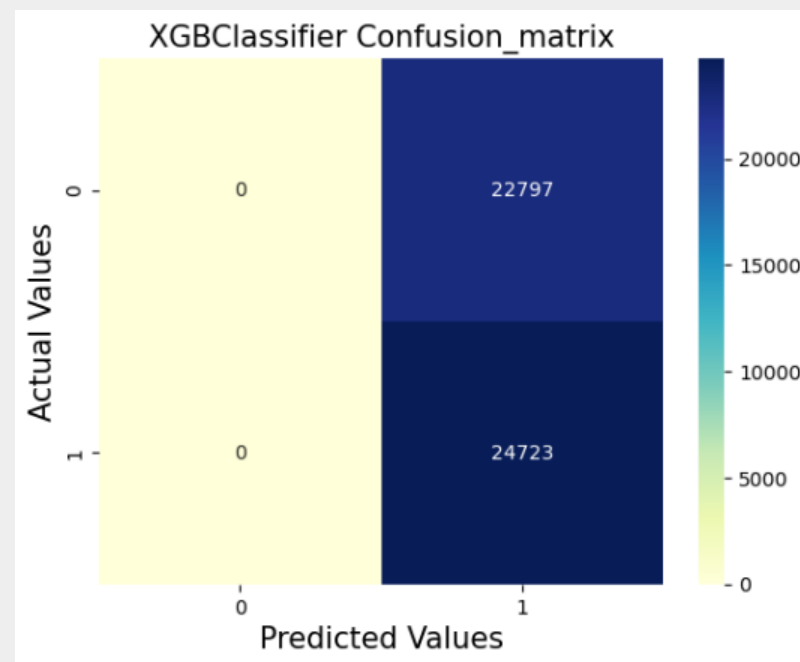
# 이슈 사항

과하게 낮은 learning\_rate(0.0001)로 인해

학습이 거의 진행되지 않거나

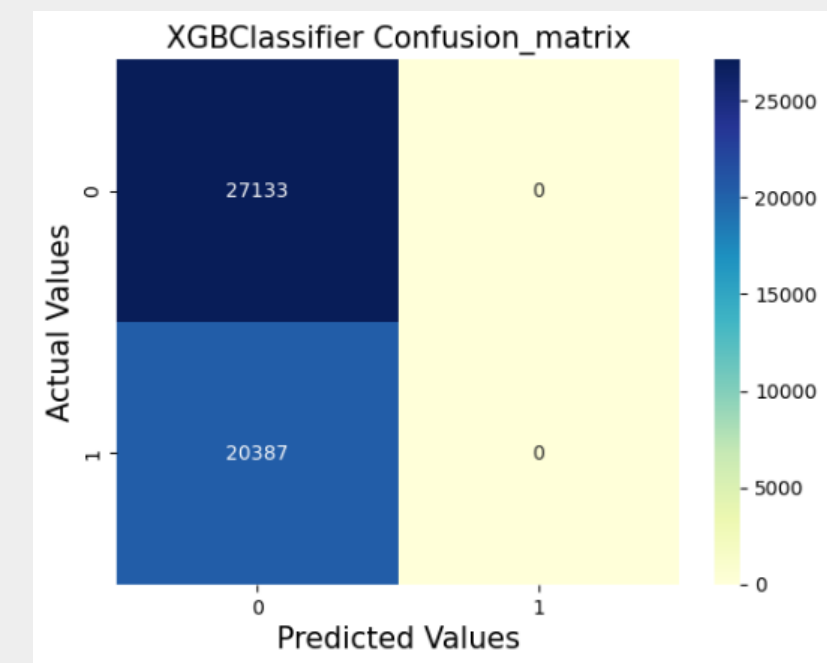
최적의 결정 경계를 찾지 못함

START - Before



오류 : 모델이 모든 데이터를 Positive로 예측  
실제 데이터는 Positive와 Negative

STOP - Before



오류 : 모델이 모든 데이터를 Negative로 예측  
실제 데이터는 Positive와 Negative

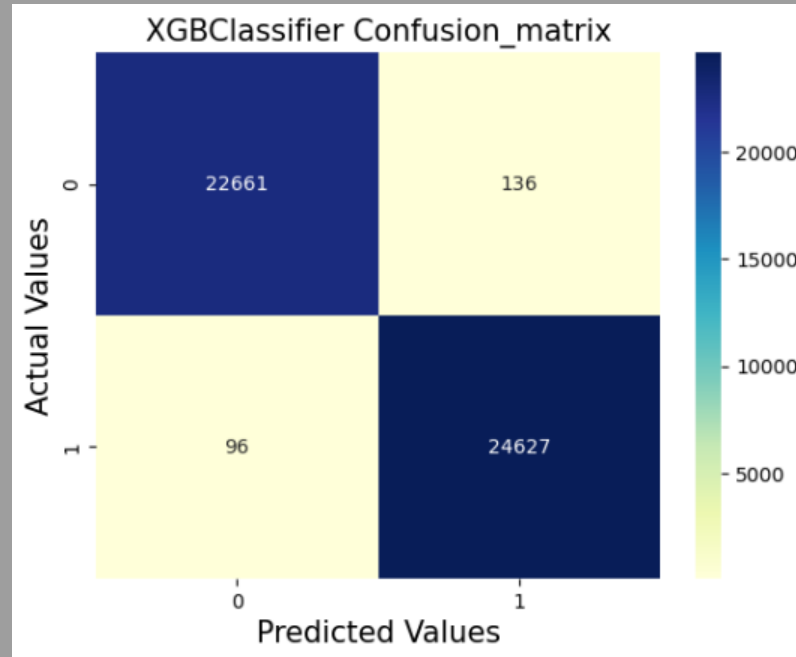
# 이슈 해결

GridSearchCV 모델을 사용해

다양한 learning\_rate[0.1, 0.01, 0.001, 0.0001] 값을 시도하고

교차 검증을 통해 최적값(0.1)을 찾음

START - After

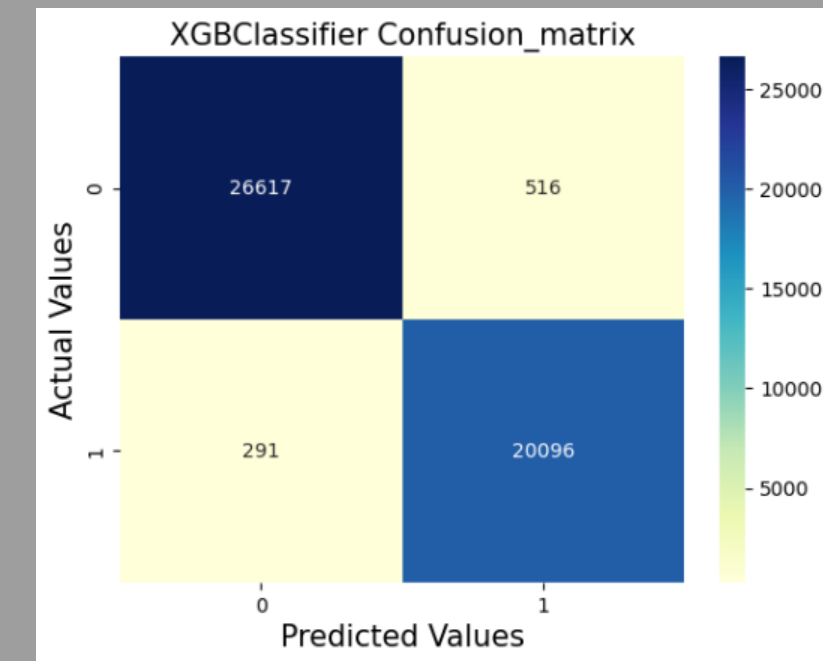


정확도 : 99.5%

정밀도 : 99.5%

재현율 : 99.6%

STOP - After



정확도 : 98.3%

정밀도 : 97.5%

재현율 : 98.6%



## 결과

- 'START' 와 'STOP' '를 각각 훈련시킨 후에는 새로운 데이터에 대해 각 모델(로지스틱, 의사결정나무, 랜덤포레스트, XGBoost)을 사용하여 시작과 종료를 예측.
- 'START' 사용한 모델의 결과와 'STOP' 사용한 모델의 결과가 각각 2개씩 나오며, 이를 종합하여 핀성형기의 시작과 종료를 식별할 수 있다.



- 과적합 처리
  - 의사결정 트리 : 트리의 깊이를 제한, 가지치기
  - 랜덤포레스트 : 트리의 갯수, 깊이 조절
  - XGBoost : learning rate 조절, 하이퍼파라미터 조절



# Thank You!

하연주  
김유림  
김희주