

2021 부산형 SW 실무인재양성

2조. 살균기 AI 데이터셋을 이용한 기계학습 및 분석



CONTENTS

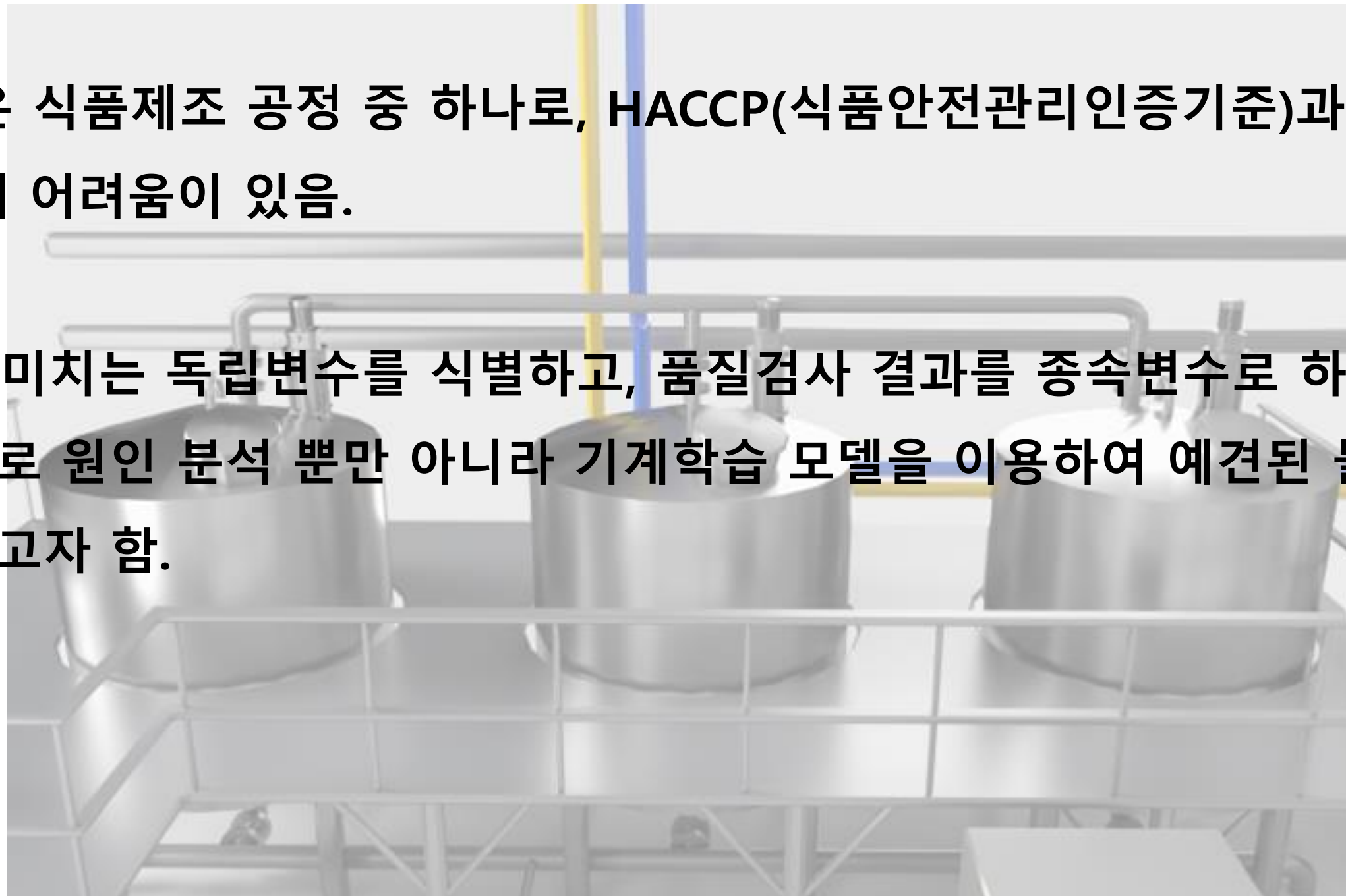
01 기획 배경

02 개발 과정

03 기능 시연(결과물)

04 결론

- 가열살균공정은 식품제조 공정 중 하나로, HACCP(식품안전관리인증기준)과 별개로 생
산품 품질 유지에 어려움이 있음.
- 품질에 영향을 미치는 독립변수를 식별하고, 품질검사 결과를 종속변수로 하여, 수집된
데이터를 기반으로 원인 분석 뿐만 아니라 기계학습 모델을 이용하여 예견된 불량률 효
율적으로 방지하고자 함.



프로젝트 팀 운영 계획	실시항목	실시기간
	각 조건별 인기메뉴 데이터 확보, 가공, 분석, 시각화	1 Day
	서버 개발(Flask)(Dash)	1 Day
	프론트엔드 개발 (HTML/CSS/Javascript)	1 Day
	데이터 분석 결과 연결	0.5 Day
수행 방법	데이터 확보: www.kamp-ai.kr (인공지능 중소벤처 제조 플랫폼)	
	데이터 분석: 파이썬(pandas, plotly)	
	웹: 파이썬(Flask, HTML, CSS, Dash)	



1. 분석 배경 및 목적

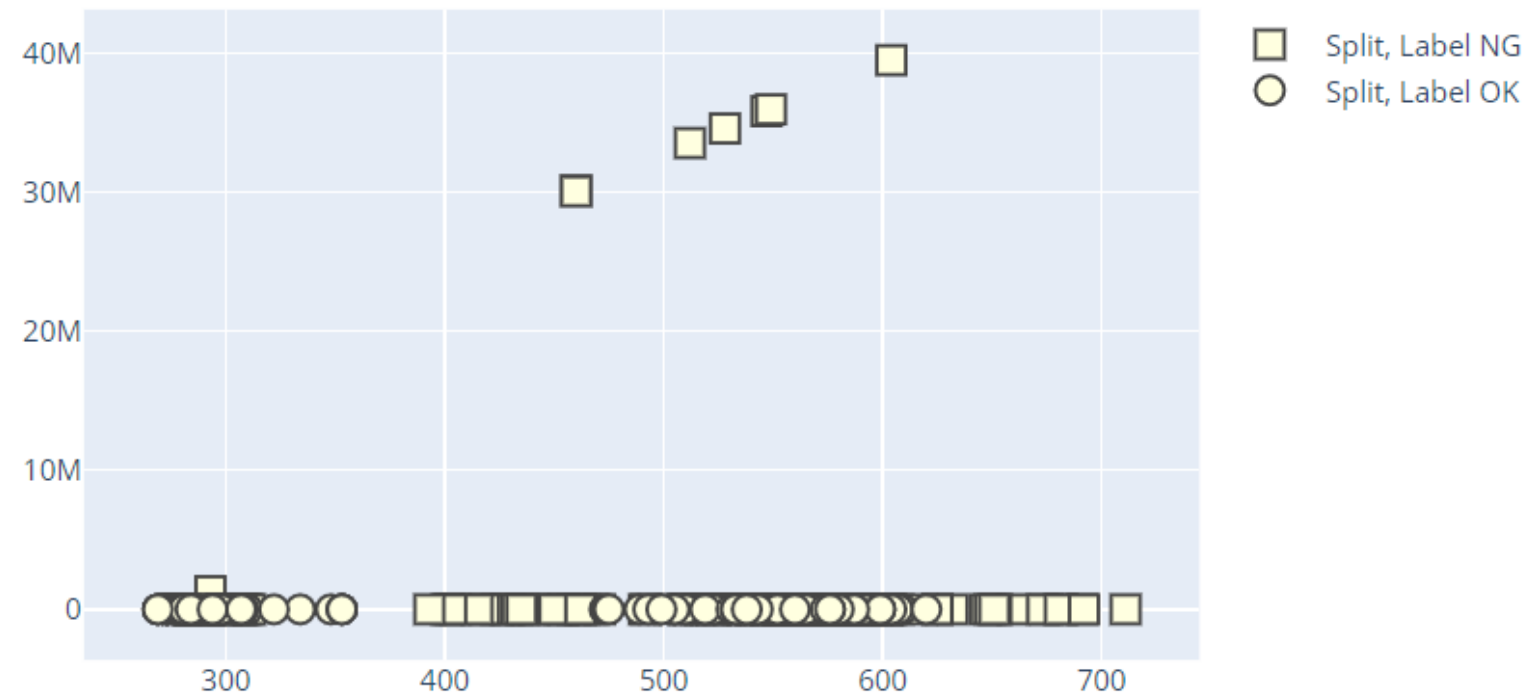
가열살균공정은 식품제조 공정 중 하나로, HACCP(식품안전관리 인증기준)과 별개로 생산품 품질 유지에 어려움이 있음.

품질에 영향을 미치는 독립변수를 식별하고, 품질검사 결과를 종속변수로 하여, 수집된 데이터를 기반으로 원인 분석 뿐만 아니라 기계학습 모델을 이용하여 예견된 불량률 효율적으로 방지하고자 함.

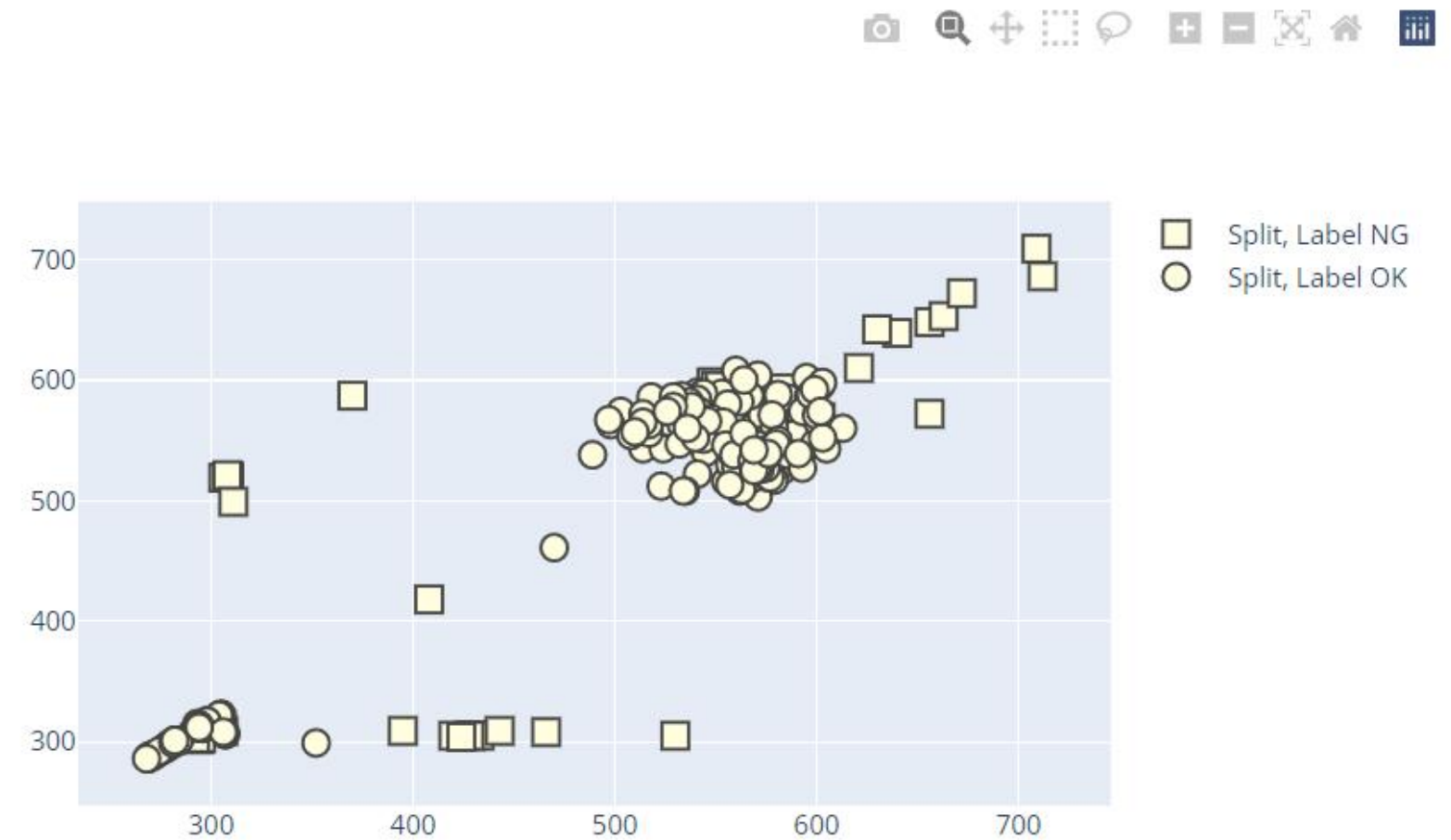
STD_DT	MIXA_PASTEUR_STATE	MIXB_PASTEUR_STATE	MIXA_PASTEUR_TEMP	MIXB_PASTEUR_TEMP	INSP
2020-03-04 6:00	1	1	551	524	OK
2020-03-04 6:30	1	1	584	536	OK
2020-03-04 7:00	1	1	584	536	OK
2020-03-04 7:30	1	1	585	536	OK
2020-03-04 8:00	1	1	585	536	OK
2020-03-04 8:30	1	1	585	536	OK
2020-03-04 9:00	1	1	585	537	OK
2020-03-04 9:30	1	1	585	538	OK
2020-03-04 10:00	1	1	585	541	OK
2020-03-04 10:30	1	1	585	543	OK
2020-03-04 11:00	1	1	585	545	OK
2020-03-04 11:30	1	1	585	548	OK
2020-03-04 12:00	1	1	584	551	OK
2020-03-04 12:30	1	1	584	554	OK
2020-03-04 13:00	1	1	582	556	OK
2020-03-04 13:15	1		551	524	NG

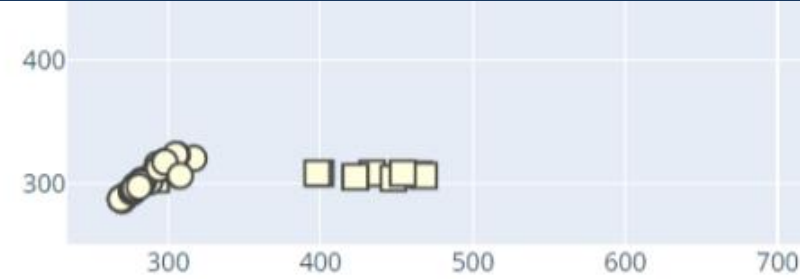
(전처리 전)

PasteurizerB 비정상값 확인 外



(전처리 후)

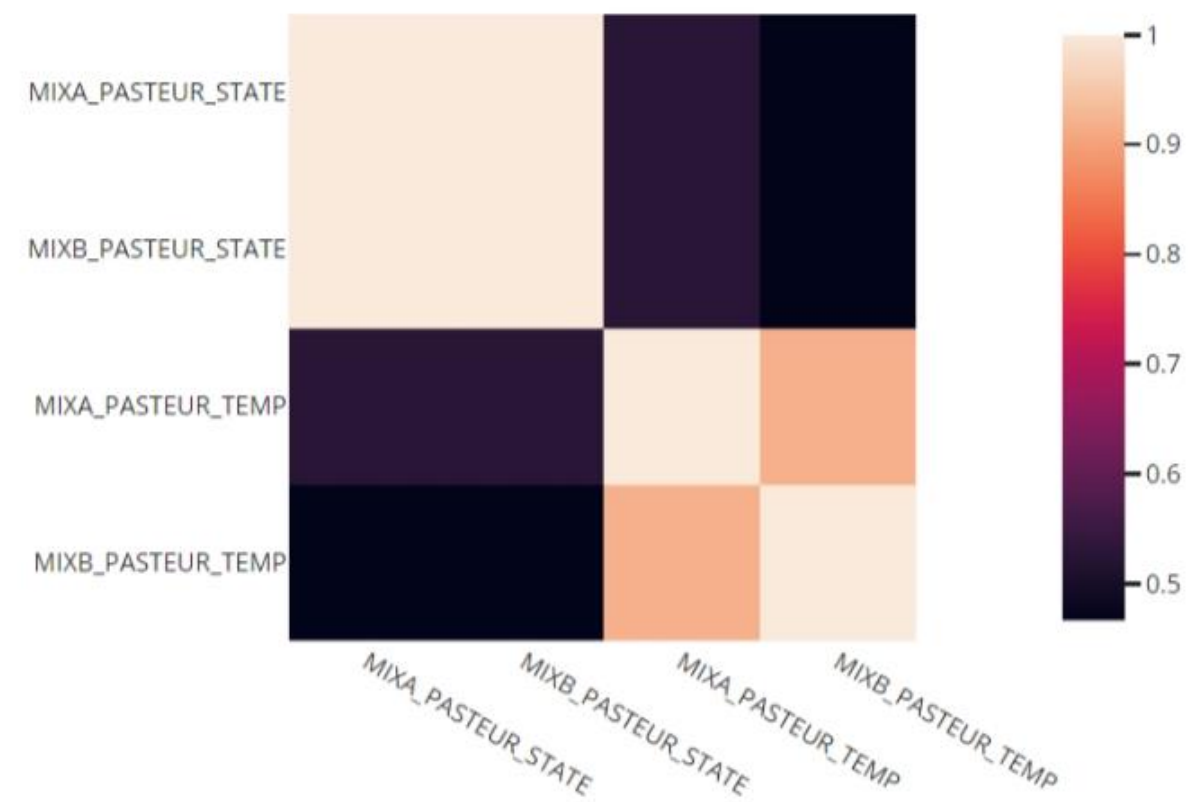




Correlation Coefficient Analysis

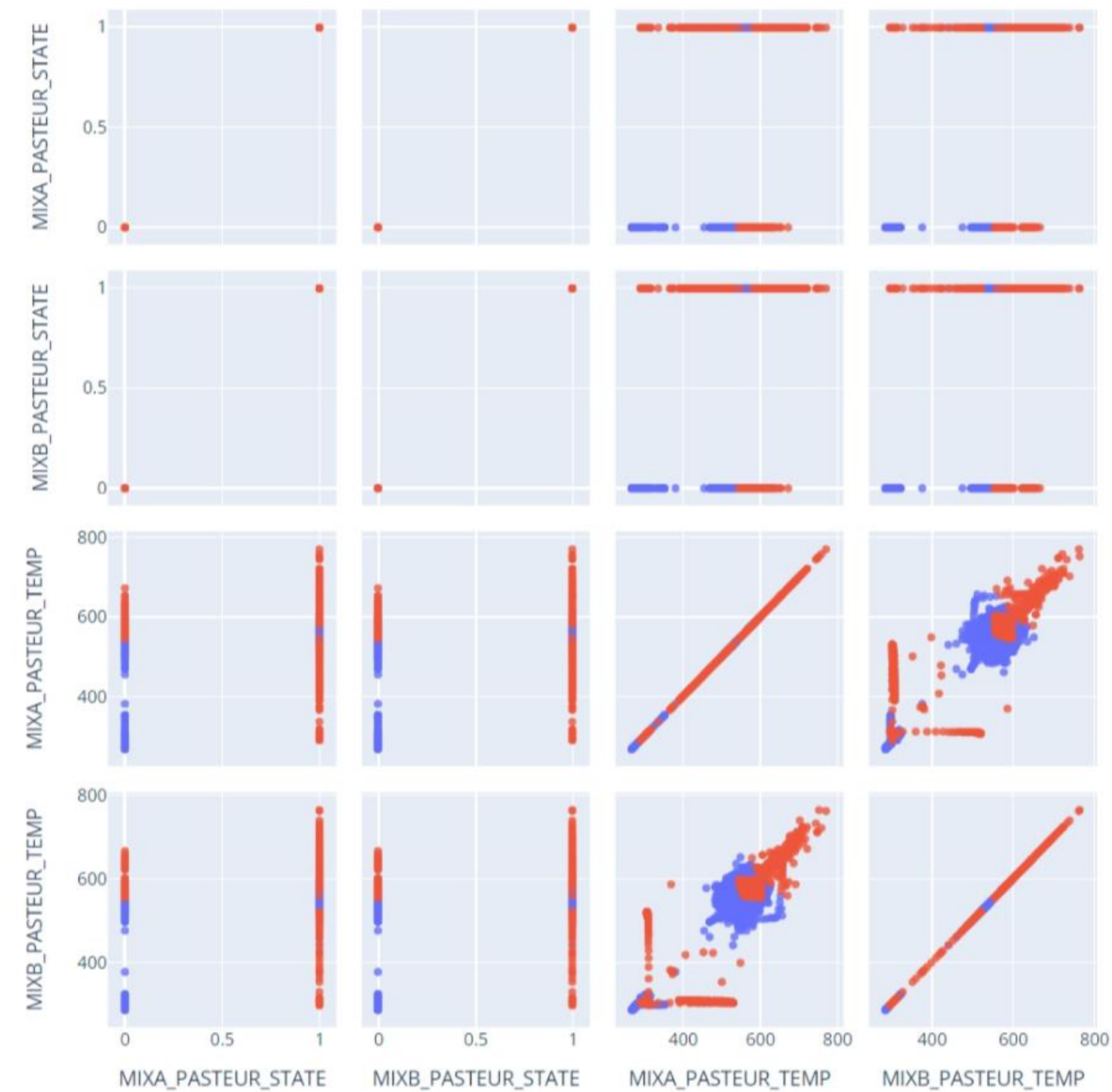
=>살균기A 온도, 살균기B 온도의 상관관계 높음

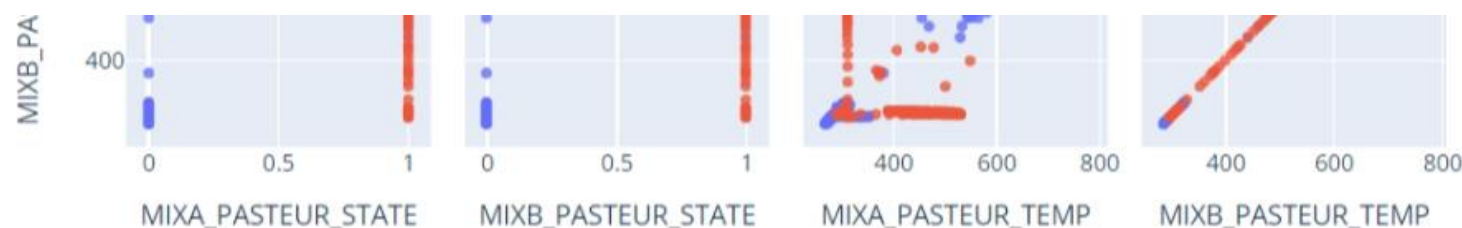
(상관계수가 높다고 하여 인과관계가 있다고 볼 수는 없음)



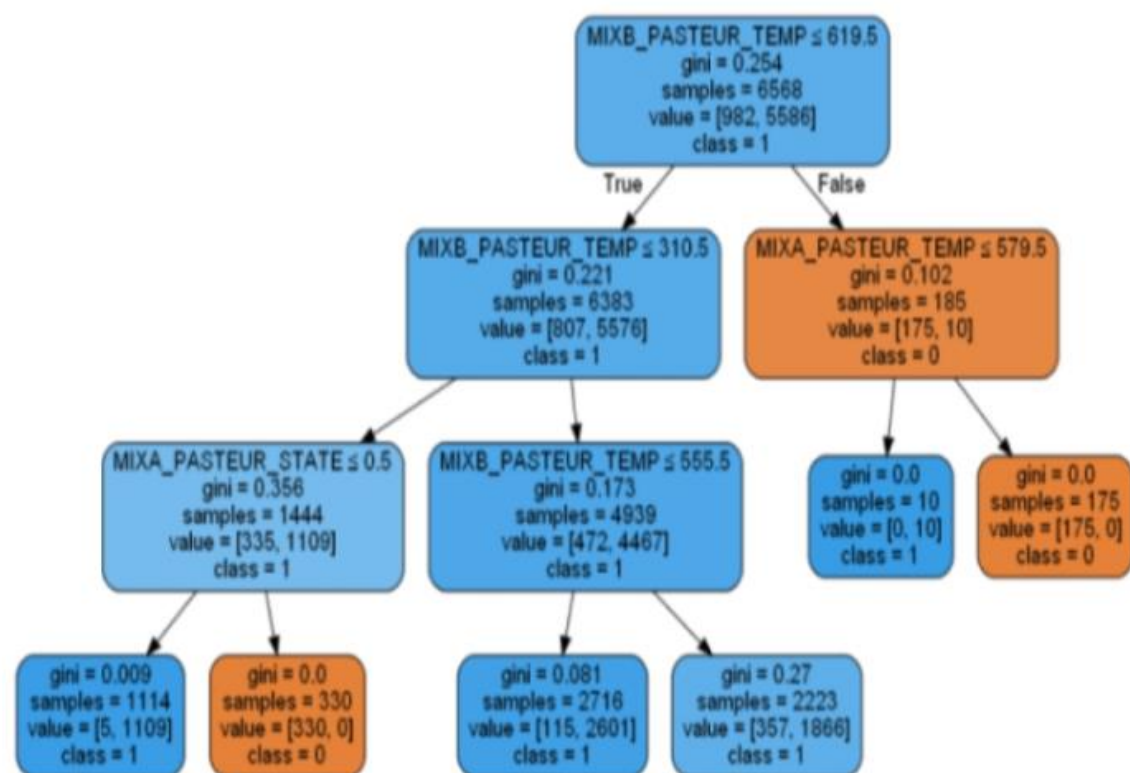
=>산점도 매트릭스

=>산점도 매트릭스





3. DecisionTree 분석



4. kNN 분석

데이터로부터 가장 가까운 k개의 다른 데이터의 레이블을 참조하는
방식의 알고리즘
(k=5)

```

roc_auc = roc_auc_score(y_test, pred)
print('오차 행렬')
print(confusion)
print(f'정확도: {accuracy: .4f}, 정밀도: {precision: .4f}, 재현율: {recall: .4f}')

get_clf_eval(y_test, dt_prediction)

```

오차 행렬
[[237 196]
 [1 2381]]
정확도: 0.9300, 정밀도: 0.9239, 재현율: 0.9996, F1: 0.9603, AUC: 0.7735

[]:

value = [5, 1100]
class = 1

value = [330, 0]
class = 0

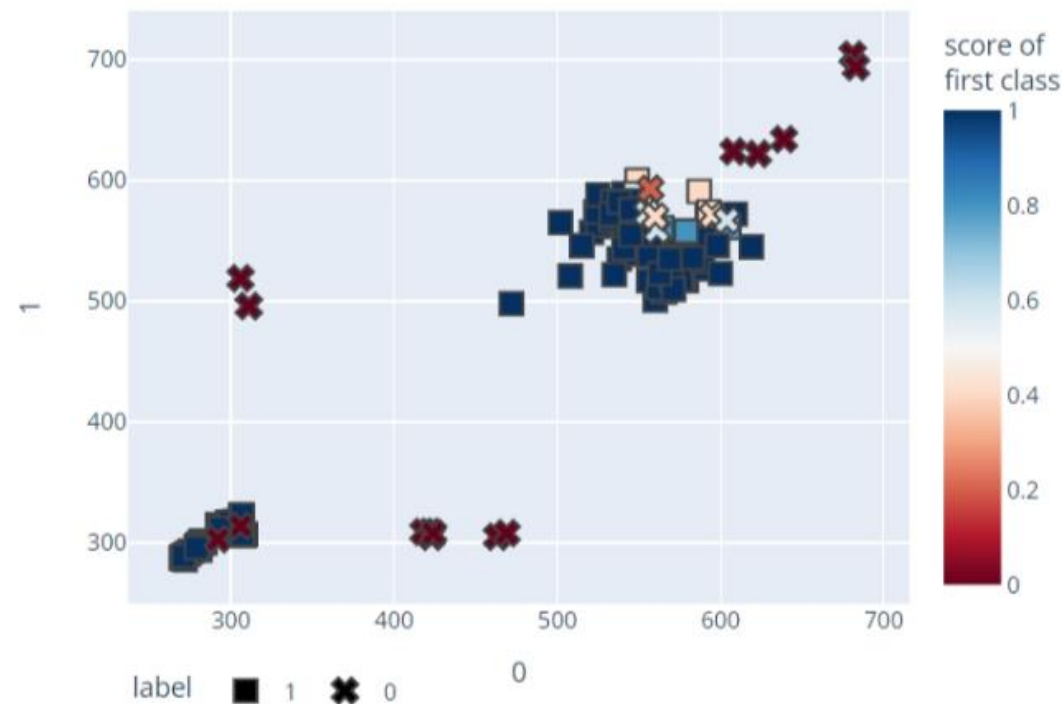
value = [110, 260]
class = 1

value = [357, 1800]
class = 1

4. kNN 분석

데이터로부터 가장 가까운 k개의 다른 데이터의 레이블을 참조하는
방식의 알고리즘

(k=5)



5. 느낀점

김태형 김현성
미니프로젝트 2조

```
PS F:\40_20211005_ML\project> f:; cd 'f:\40_20211005_ML\project'; & 'f:\40_20211005_ML\python\python_2021_10_1265161270\pythonFiles\lib\python\debugpy\launcher' '6128'
오차 행렬
[[ 364  60]
 [  32 2357]]
정확도: 0.9673, 정밀도: 0.9752, 재현율: 0.9866, F1: 0.9809, AUC: 0.9225
PS F:\40_20211005_ML\project>
```

1. 학습모델별 결과 (Decision Tree, kNN)
-> Decision Tree보다 kNN 모델이 우수
(3%p↑)
2. MIXB_PASTEUR_TEMP가 619.5를 초과하고,
MIXA_PASTEUR_TEMP가 579.5 이하인 경우,
양품으로 분류
3. MIXB_PASTEUR_TEMP가 619.5를 초과하고,
MIXA_PASTEUR_TEMP가 579.5 초과인 경우,
불량으로 분류
4. MIXB_PASTEUR_TEMP가 619.5를 이하인 경우,
양품으로 분류*

*(조건에 의해 양품여부가 달라질 수도 있음)

No part of this publication may be circulated / quoted / or reproduced
for distribution outside the client organization without prior written approval from KSA

감사합니다.

KSA 한국표준협회
KOREAN STANDARDS ASSOCIATION

데이터셋 출처

(<https://www.kamp-ai.kr/front/faq/faq-list.jsp> 에 의거 표기)

중소벤처기업부,
Korea AI Manufacturing Platform(KAMP),
살균기 AI 데이터셋,
KAIST (수행기관: (주) 임픽스/한양대학교 산학협력단/(주)아큐라소프트),
2020.12.14 (최종수정일: 2021.02.16),
<https://kamp-ai.kr>