

제조 공정 생산성 향상을 위한 시각화 및 머신러닝 기반 예측 모델 개발

AI 제조 4팀

김남훈, 김정민, 이상현, 이슬인, 임규호, 윤종현, 정용규

Contents

01 팀 구성 및 역할

02 프로젝트 개요

- 문제 정의
- 데이터 설명
- 데이터 문제

03 프로젝트 수행 및 과정

- 모델 설명
- 모델 정확도 비교
- 피드백 적용

04 프로젝트 수행 결과

- 기대효과
- 활용방안

팀 구성 및 역할

< 팀장 >



김정민

20학번 소프트웨어전공

역할 : PPT 제작, EDA

< 팀원 >



김남훈

18학번 IT경영

역할 : DDN 모델링



윤종현

17학번 IT경영

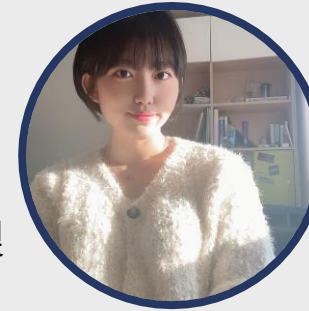
역할 : EDA, KNN 모델링 및
검증, 평가



이상헌

18학번 게임공학

역할 : 랜덤포레스트 및
의사결정나무



이슬인

19학번 전자공학

역할 : PPT 보조, DDN



임규호

16학번 생명화학공학

역할 : 자료조사



정용규

19학번 컴퓨터공학

역할 : 랜덤포레스트 및
의사결정나무

02 프로젝트 개요



저온 살균

살균의 한 방법,

상압에서 100℃ 이하의 온도에서 진행되는 살균 방식

02 프로젝트 개요



살균 공정은 원재료의 전처리를 수행하는 두 번째 단계로, 살균 품질에 따라 후공정 및 완제품 품질에 미치는 영향이 큼

살균 공정은 HACCP 주요 관리점이기 때문에 식품안전관리인증 이행 측면에서도 중요한 공정

02 프로젝트 개요

A 유업

분말 유크림, 기능성 조제
분말 등을 생산하는
중소제조업체

- 살균 공정은 일반적으로 **살균 온도와 살균 시간**을 지표로 CCP 이행여부를 판단함
- 제조현장에서는 기존 자료와 시험 결과 등을 바탕으로 **CCP 기준을 준수**하면서 동시에 **목표 품질을 위한 공정 기준값을 설정**해 공정을 수행함
- 가열 방식의 특성 상 살균 공정동안 내용물의 온도는 일정하게 유지되지 않고 계속해서 변동
- 설정온도가 유지되도록 이에 맞춰 가열 온도를 변경하지만, 온도가 조정되는 동안의 시차가 있어 **어느 정도의 온도 변화는 불가피**함

02 문제점 제시

대기업의 경우,
충분한 자본과 고도화된 기술로
데이터에 기반한 살균기의
온도 모니터링 기술을 통해
자동 온도 조절 시스템을 사용



상대적으로 자본과 기술력이 부족한
중소제조기업의 경우,

사람의 주관적 판단에 따른
온도 설정으로 인해
원활하고 일관적인 제어가 되지않음

제품의 품질과 CCP 평가기준을
지키는 것이 쉽지 않음

02 해결 방안



작업자의 주관적 판단과 노후화된 살균기의 자체 기능에만 의존하는 방법이 아닌,
데이터에 기반한 품질예측 방법이 필요함

따라서, 살균 여부의 주요 지표인 살균온도와 제품 불량 데이터의 머신러닝을 통해
실시간 품질예측 시스템을 만들고 공정제어에 이용할 수 있도록 함

02 데이터 설명

A	B	C	D	E	F
STD_DT	MIXA_PAS	MIXB_PAS	MIXA_PAS	MIXB_PAS	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
2020/03/04 13:30	1	1	580	558	OK
2020/03/04 14:00	1	1	577	561	OK
2020/03/04 14:30	1	1	573	564	OK
2020/03/04 15:00	1	1	572	566	OK
2020/03/04 15:30	1	1	566	569	OK
2020/03/04 16:00	1	1	559	570	OK
2020/03/04 16:30	1	1	554	571	OK

변수명	설명	데이터 타입
STD_DT	날짜, 시간 (YYYY-MM-DD HH : MM : SS)	문자
MIXA_PASTEUR_STATE	살균기 A 가동상태	숫자
MIXB_PASTEUR_STATE	살균기 B 가동상태	숫자
MIXA_PASTEUR_TEMP	살균기 A 살균온도	숫자
MIXB_PASTEUR_TEMP	살균기 B 살균온도	숫자
INSP	불량여부	문자

- _PASTEUR_STATE 값 0 정지 1 운전
- _PASTEUR_TEMP 소수점 1자리가 생략되어있음(ex. 456 > 45.6℃)
- INSP OK : 양품, NG : 불량
- 약 8.5개월 간의 수집한 데이터를 나타내는 csv 파일

02 데이터 문제

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 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
2020/03/04 13:30	1	1	580	558	OK
2020/03/04 14:00	1	1	577	561	OK
2020/03/04 14:30	1	1	573	564	OK

2020/03/05 16:15			584	536	OK
2020/03/05 16:16			585	536	OK
2020/03/05 16:17			585	536	OK
2020/03/05 16:18			585	536	OK
2020/03/05 16:19			585	537	OK
2020/03/05 16:20			585	538	OK

데이터를 측정한 시간이 일정하지 않음



30, 15, 1분 단위로 측정된 데이터를 가공을 통해
하나의 통일된 기준으로 자료를 생성하는 게
적합하다고 판단함

02 데이터 문제

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

2020/03/05 16:14	1		584	536	OK
2020/03/05 16:15			584	536	OK
2020/03/05 16:16			585	536	OK
2020/03/05 16:17			585	536	OK
2020/03/05 16:18			585	536	OK
2020/03/05 16:19			585	537	OK

2020/03/05 19:18			568		OK
2020/03/05 19:19			569		OK
2020/03/05 19:20			569	542	OK
2020/03/05 19:21			570		OK
2020/03/05 19:22			570		OK
2020/03/05 19:23			571	541	OK

데이터에서 **상태 결측치**와 **온도 결측치**가 존재함

양품인지 불량품인지 판단할 수 있는 기준을 파악하기 모호함

03 EDA

2003/12/20 22:00	1	1	576	537	OK
2003/12/20 22:30	1	1	576	537	OK
2003/12/20 23:00	1	1	574	537	OK
2003/12/20 23:30	1	1	571	537	OK
3/13/20 0:00	1	1	568	540	OK
3/13/20 0:30	1	1	565	544	OK
3/13/20 1:00	1	1	560	548	OK
3/13/20 1:30	1	1	557	550	OK
3/13/20 2:00	1	1	553	555	OK
3/13/20 2:30	1	1	550	557	OK

EDA 결과 시간 표기가 다른 걸 확인함.

- 8.5 개월의 데이터이기 때문에
온도 변화의 경향성을 보려
30분단위로 통합하기로 결정함.

-> **전처리** 후 사진과 같이 **날짜표기법이 바뀌어있거**
나,
날짜는 같으나 온도가 다른 **중복되는 데이터가**
발생하는 걸 발견함.

03 EDA

210666	2020/11/11 16:30			595	562	OK
210696	2020/11/11 17:00			553		OK
210726	2020/11/11 17:30			592	554	OK
210755	2020/11/11 18:00			559	595	OK
210785	2020/11/11 18:30			590	553	OK
209742	2020/11/11 1:00				561	OK
209772	2020/11/11 1:30			556	593	OK
209802	2020/11/11 2:00			603	553	NG
209832	2020/11/11 2:30			559	598	OK
209862	2020/11/11 3:00			606		NG
209892	2020/11/11 3:30			556	600	OK

EDA 결과 결측치의 경우, 두 가지로 나뉨

1. 상태 결측치

- 1로 시작하면 다음 숫자가 나오기 전까지 같은 숫자로 채워넣기로 결정
- 표기할 때 1과 0을 한 번씩만 쓰고 일정기간 동안 표기가 없는 경향성을 발견해 표기를 생략했다고 판단함

2. 온도 결측치

- 선형보간법을 이용해 앞과 뒤의 온도 평균값으로 채워넣기로 결정함.

03 EDA

2020/03/12 22:00	1	1	576	537	OK
2020/03/12 22:30	1	1	576	537	OK
2020/03/12 23:00	1	1	574	537	OK
2020/03/12 23:30	1	1	571	537	OK
2020/03/13 0:00	1	1	568	540	OK
2020/03/13 0:30	1	1	565	544	OK
2020/03/13 1:00	1	1	560	548	OK
2020/03/13 1:30	1	1	557	550	OK
2020/03/13 2:00	1	1	553	555	OK
2020/03/13 2:30	1	1	550	557	OK
2020/03/13 3:00	1	1	548	560	OK

날짜표기법 처리

2020/11/07 0:00	1	1	650	637	OK
2020/11/07 0:30	1	1	612	600	NG
2020/11/07 1:00	1	1	692	644	OK
2020/11/07 1:30	1	1	683	669	OK
2020/11/07 2:00	1	1	669	702	OK
2020/11/07 2:30	1	1	682	694	OK
2020/11/07 3:00	1	1	696	656	OK
2020/11/07 3:30	1	1	688	686	OK
2020/11/07 4:00	1	1	647	643	OK
2020/11/07 4:30	1	1	620	614	NG
2020/11/07 5:00	1	1	596	590	OK

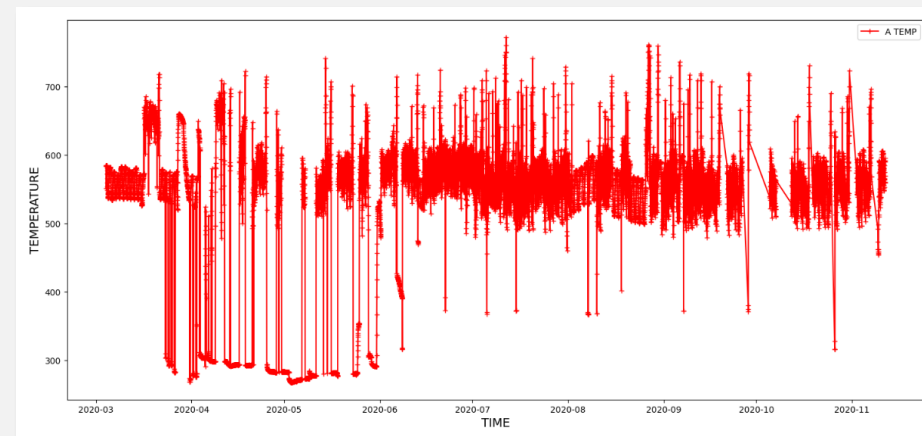
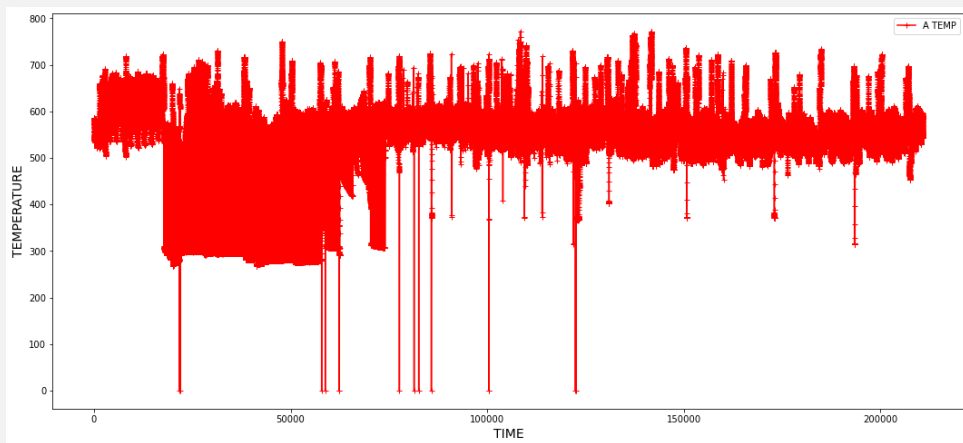
결측치 처리

03 EDA

```
OK      7454
NG      2882
Name: INSP, dtype: int64
```

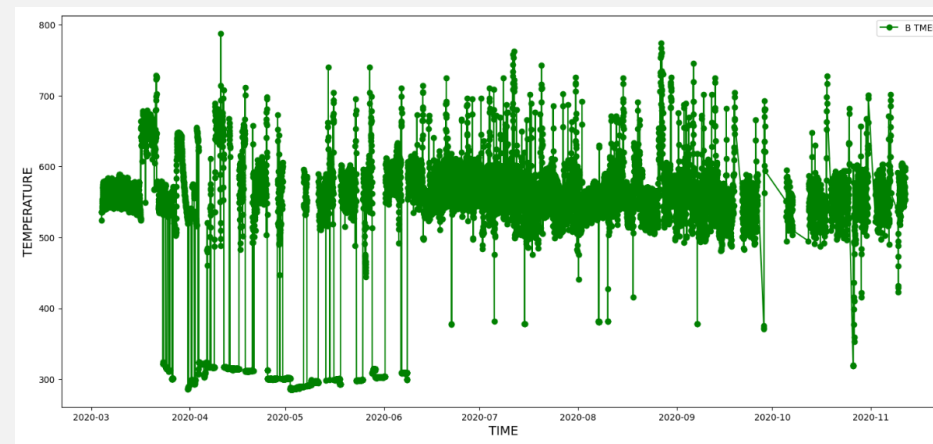
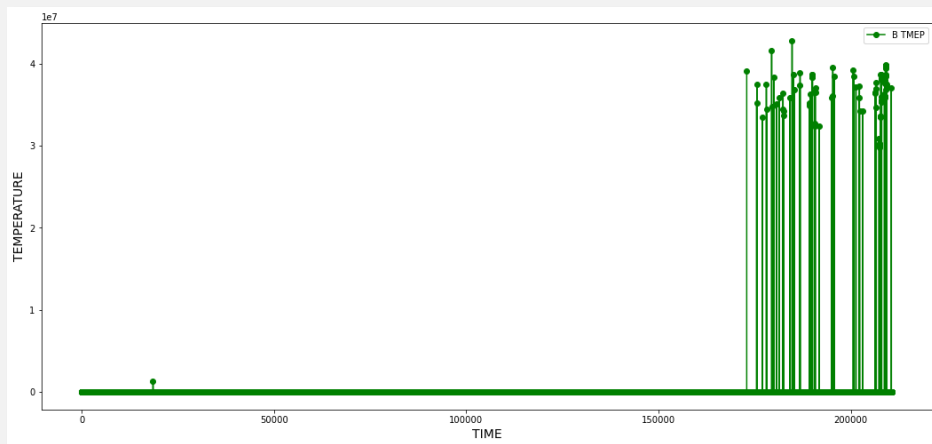
데이터 전처리 후
양품과 불량품의 개수를
나타낸 표이다.

03 EDA



온도 A의 데이터 전처리 전후 과정
선형보간법을 적용한 뒤
온도변화가 어떻게 되는지 시각화한 모습이다.

03 EDA



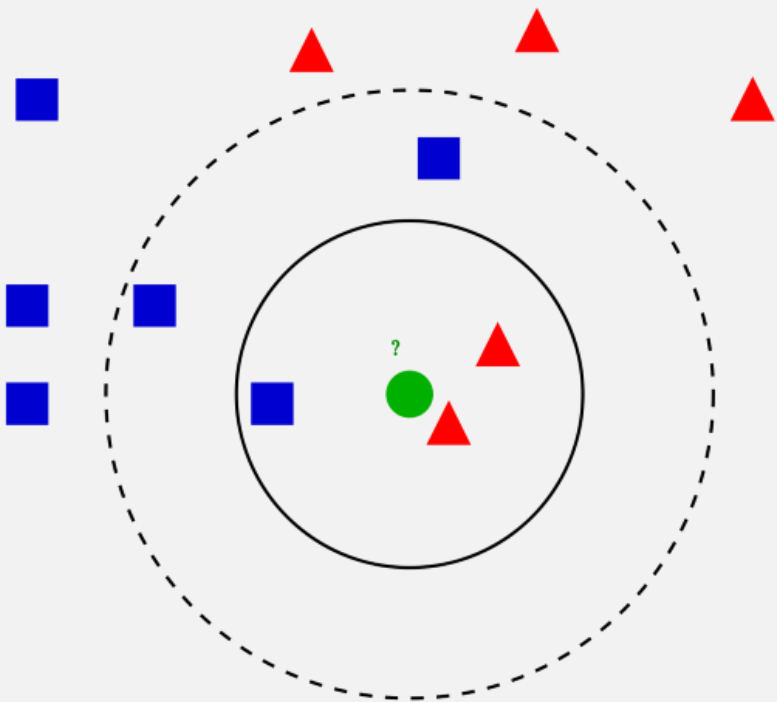
온도 B의 데이터 전처리 전후 과정
선형보간법을 적용한 뒤
온도변화가 어떻게 되는지 시각화한 모습이다.

03 모델 정의

데이터를 분석해 **패턴을 발견**하거나 **예측을 수행**하는 알고리즘의 표현 방법
머신러닝 모델의 종류는 다음과 같음

ML 모델 유형	활용사례
선형 회귀/분류	수치 데이터의 패턴
그래픽 모델	부정 감지 또는 감정 인식
의사결정트리/랜덤 포레스트	결과 예측
딥러닝 신경망	컴퓨터 비전, 자연어 처리 등

03 모델 - KNN



KNN – k-Nearest Neighbors Model

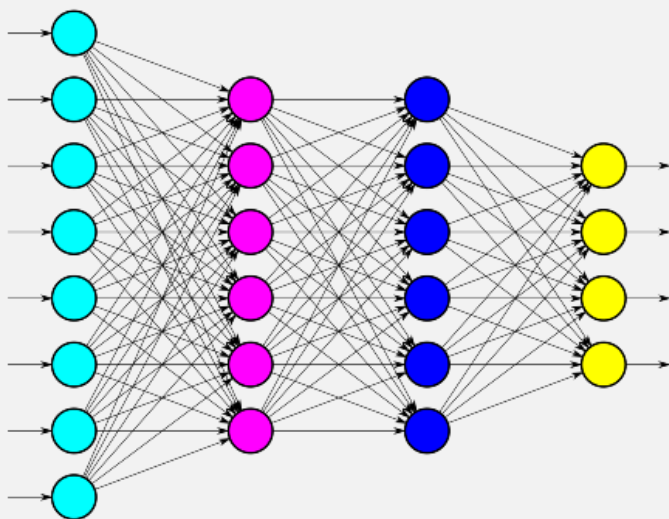
분류 문제에 사용하는 알고리즘
새로운 데이터가 들어왔을 때
기존 데이터의 그룹 중 어떤 그룹에 속하는지
분류하는 역할을 한다.

주로 이미지 분류 문제에 사용됨

탐색할 이웃 수 k 와 거리 측정 방법

두 가지를 파라미터로 입력받아 분류하는 방식

03 모델 - DNN



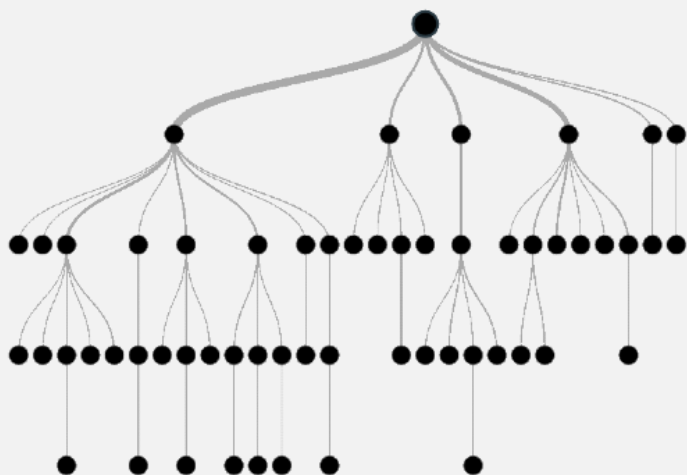
DNN – Deep Neural Network Model

사람의 신경망 원리와 구조를 모방해 만든
기계학습 알고리즘

인공신경망에서의 자극, 즉 Input data가
임계값을 넘어가면 여러 층에 거쳐 Output 데이터로 변
환

컴퓨터 스스로 분류레이블을 만들어내 공간을 왜곡하고
데이터를 구분짓는 과정을 반복해 최적의 결과 도출

03 모델 – Random Forest



Random Forest

일련의 분류 규칙을 통해 데이터를 분류, 회귀하는 지도 학습 모델

특정 기준에 따라 데이터를 구분
분기 기준을 선택하기 위해 불순도를 사용

- 불순도란?

- 해당 범주 안에 서로 다른 데이터가 얼마나 섞여있는지를 판별할 수 있는 개념
- 다양한 개체들이 섞여있을 수록 높아짐
- 불순도는 말단노드로 갈 수록 감소되어야함

03 해당 모델 선택 이유

모델	특징
KNN	<ol style="list-style-type: none">1. 기존 분류 체계 값을 모두 검사해 비교, 높은 정확도2. 오류 데이터 비교대상에서 제외
DNN	<ol style="list-style-type: none">1. 연속형, 범주형 변수에 상관없이 모두 분석 가능2. 예측력이 다른 머신러닝 기법들에 비해 상대적으로 우수함
Random Forest	<ol style="list-style-type: none">1. 통계모델에 요구되는 가정에 자유로움2. 직관적인 해석 가능

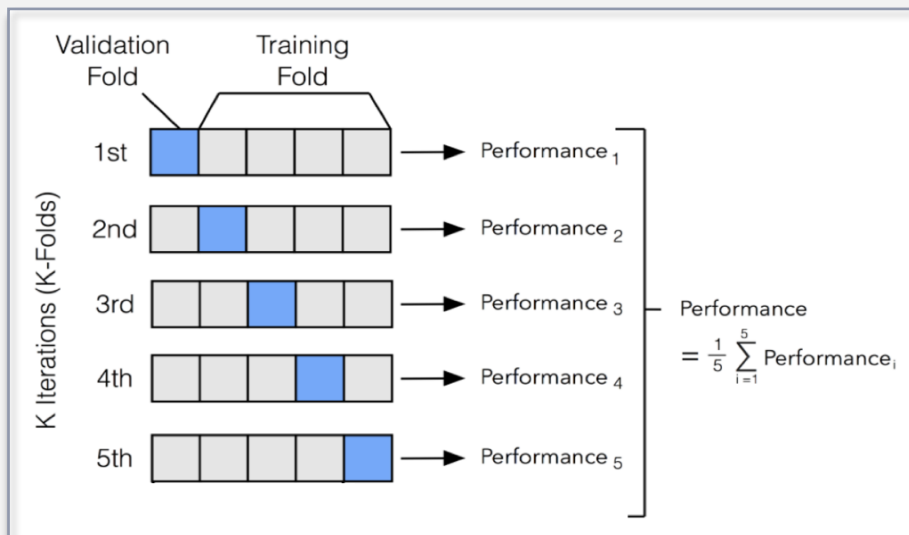
03 모델링 시 공통사항

모델	공통사항
KNN	1. Test size = 0.25 로 고정 2. Stratify = df['INSP'] 로, 원본 데이터 비율 유지하면서 Train_test_split 진행 3. 선형 보간법 을 이용해 결측치를 채워넣은 데이터 사용
DNN	
Random Forest	

! *Train_test_split*이란?

- 모델링을 위해 훈련 데이터 세트와 테스트 데이터 세트를 나눌 때 사용하는 *sklearn*의 기능 중 하나

03 Cross Validation

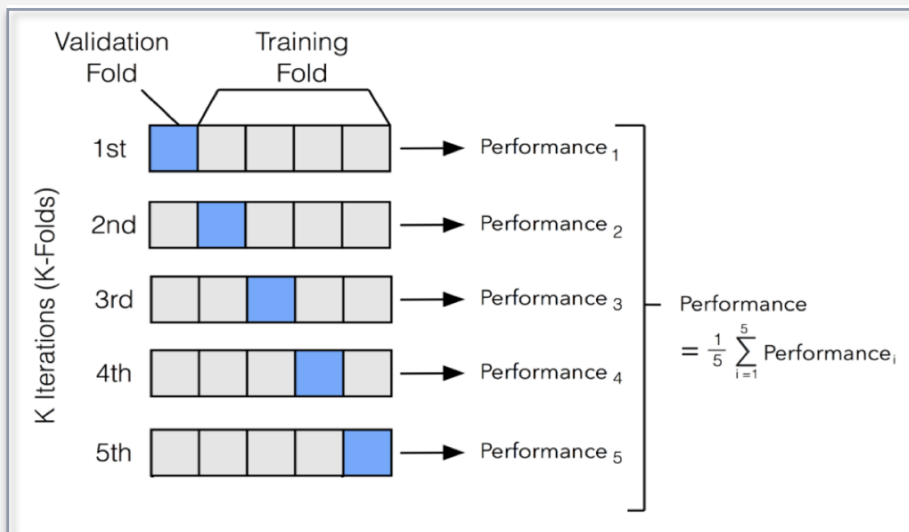


교차 검증이란?

데이터를 여러 번 반복해서 나누고
여러 모델을 학습해 성능을 평가하는 방법

학습 데이터, 평가용 데이터 세트의 비율을 변화시키
며,
일반화된 성능을 얻기 위해 진행

03 k-fold Cross Validation



K-겹 교차 검증이란?

교차 방법 중 하나

데이터를 **k개로 분할** 한 뒤, k-1개를 학습용 데이터 세트, 1개를 평가용 데이터 세트로 사용하며

이를 **k번 반복**해 **평균적인 성능 지표**를 얻어내는 방법

각 모델별 공통사항

K값 10으로 고정해 교차검증 수행함

03 모델별 정확도 비교

Confusion Matrix			
		Predicted	
		Positive	Negative
Actual	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

모델을 평가하는 지표에는 여러가지가 존재함

1. F1 Score
2. ROC curve
3. Confusion Matrix

03 모델별 정확도 비교

Confusion Matrix			
		Predicted	
		Positive	Negative
Actual	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

정확도 - 전체 표본 중 정확히 분류된 표본의 수의 비율
 $(T.P + T.N) / (T.P + F.N + F.P + T.N)$

- 예측의 성공에만 초점을 두기 때문에
편향성에 따라 달라질 수 있으므로 정확도 만으

로는

정밀도 - 모델을 제대로 평가할 수 없음
모델이 정답이라고 예측한 것 중에서 실제로 정답
인

비율

$$(T.P) / (T.P + F.P)$$

재현율 - 실제 정답 중에서 모델이 정답을 맞춘 비율
 $(T.P) / (T.P + F.N)$

03 모델별 정확도 비교

Confusion Matrix			
Actual	Positive	(True Positive)	(False Negative)
	Negative	(False Positive)	(True Negative)

그러나 재현율과 정확도만으로는
모델을 제대로 평가할 수 없음

Hit rate –

재현율과 정확도는 모델이 True라고
예측한 것에 초점을 맞추고 있음

-> 확실하게 분류할 수 있는 거만 예측해 정확도를 높
인다면 이상적인 모델이 만들어지지 않을 것

03 모델별 정확도 비교

$$F1 \text{ Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

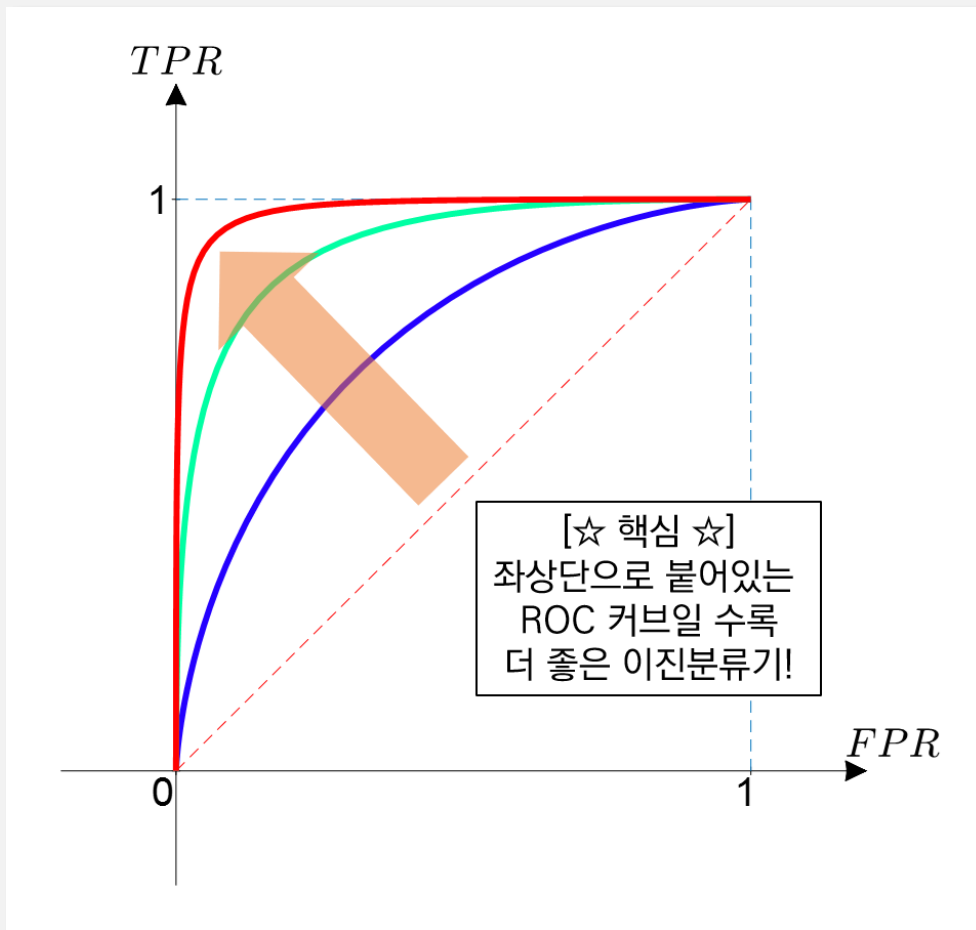
그러나 재현율과 정확도만으로는 모델을 제대로 평가할 수 없음

	Actual Positive	Actual Negative
Predicted Positive	TP (True Positive)	FP (False Positive)
Predicted Negative	FN (False Negative)	TN (True Negative)

따라서, 두 가지를 고려한다면 실제 양품과 모델이 양품이라고 예측한 비율을 함께 고려하게 되어 제대로 평가할 수 있음.

따라서, 재현율과 정확도의 관계를 살펴볼 수 있는 F1-score를 이용해 모델을 평가한다

03 모델별 정확도 비교



ROC 커브란?

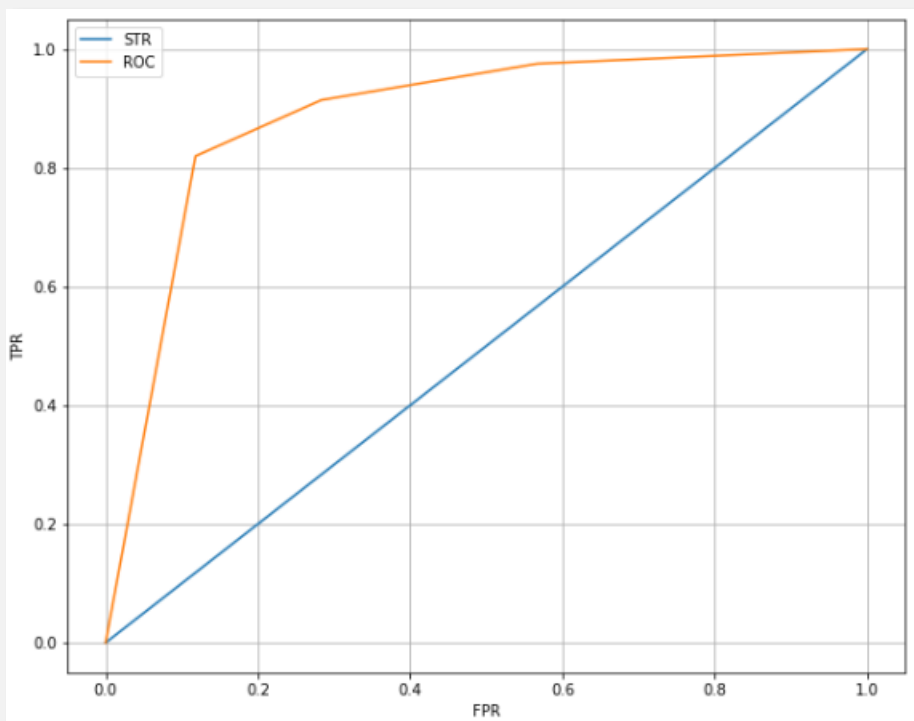
F1 Score와 유사하게 재현율과 특이도의 관계를 나타내는 그래프

특이도 : 모델이 거짓이라고 분류한 것에서 실제로 거짓인 비율

모델의 판별력을 살펴보기 위한 평가지표 중 하나이다.

좌상단으로 붙어있는 ROC일 수록 성능이 좋은 모델이다.

03 KNN



k	실행 시 정확도
1	0.86266925
2	0.84332689
3	0.86073501
4	0.86847195
5	0.87234043
6	0.87330754
7	0.85866409
8	0.85188771
9	0.86447241
10	0.85382381

- 선형보간법 방법을 적용했을 때의 ROC 커브를 출력한 결과
- 10 겹 교차검증을 수행했을 때의 정확도

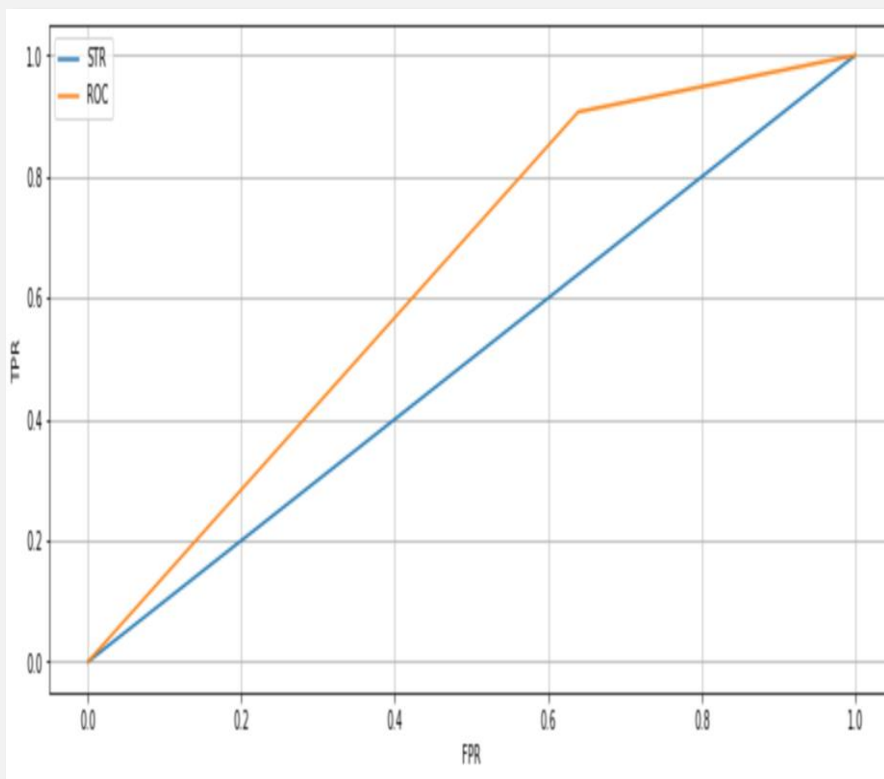
03 KNN

	Precision	Recall	f1-score	Support
0	0.77	0.75	0.76	720
1	0.91	0.91	0.91	1864
정확도			0.87	2584
Macro avg	0.84	0.83	0.84	2584
Weighted avg	0.87	0.87	0.87	2584

Classification_report 를 이용해 모델 평가지표를 출력한 결과

Macro avg, Weighted avg가 비슷할 수록
균형이 잘 잡힌 데이터, 즉 좋은 모델이라고 볼 수 있음.

03 DNN



k	실행 시 정확도
1	0.72380954
2	0.75238096
3	0.74642854
4	0.75
5	0.78690475
6	0.75595235
7	0.74642854
8	0.78214287
9	0.78069132
10	0.74255067

- 선형보간법 방법을 적용했을 때의 ROC 커브를 출력한 결과
- 10 겹 교차검증을 수행했을 때의 정확도
- 다른 모델에 비해 상대적으로 정확도가 낮음

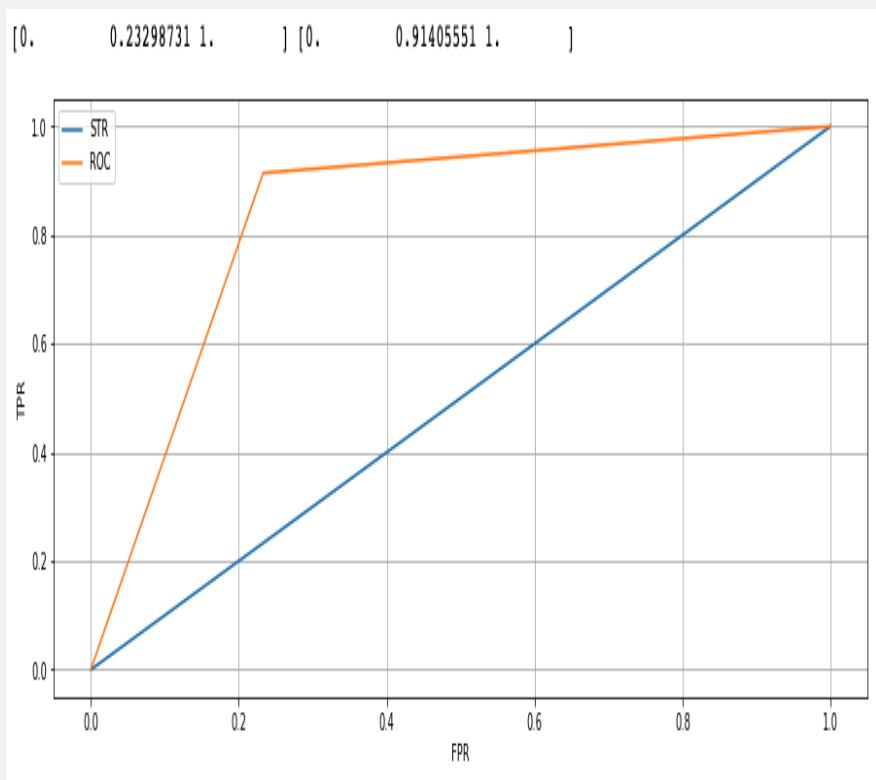
03 DNN

	Precision	Recall	f1-score	Support
0	0.60	0.36	0.65	720
1	0.79	0.91	0.84	1864
정확도			0.75	2584
Macro avg	0.69	0.63	0.65	2584
Weighted avg	0.73	0.75	0.73	2584

Classification_report 를 이용해 모델 평가지표를 출력한 결과

Macro avg, Weighted avg가 비슷할 수록
균형이 잘 잡힌 데이터, 즉 좋은 모델이라고 볼 수 있음.

03 Random Forest



k	실행 시 정확도
1	0.87137330
2	0.86847195
3	0.87620889
4	0.87524177
5	0.8810444
6	0.8713733
7	0.8780251
8	0.8576960
9	0.8693126
10	0.8780251

- 선형보간법 방법을 적용했을 때의 ROC 커브를 출력한 결과
- 10 겹 교차검증을 수행했을 때의 정확도

03 (3) Random Forest

	Precision	Recall	f1-score	Support
0	0.78	0.77	0.77	867
1	0.91	0.91	0.91	2234
정확도			0.87	3101
Macro avg	0.84	0.84	0.84	3101
Weighted avg	0.87	0.87	0.87	3101

Classification_report 를 이용해 모델 평가지표를 출력한 결과

Macro avg, Weighted avg가 비슷할 수록
편향되지 않은 모델이라고 볼 수 있음.

03 결과분석

OK	Precision	Recall	f1-score	Support
0	0.77	0.75	0.76	720
1	0.91	0.91	0.91	1864

OK	Precision	Recall	f1-score	Support
0	0.60	0.36	0.65	720
1	0.79	0.91	0.84	1864

OK	Precision	Recall	f1-score	Support
0	0.78	0.77	0.77	867
1	0.91	0.91	0.91	2234

0 : 불량품 1 : 양품

양품을 판별하는 건 높은 데에 비해
불량품을 판별하는 건 낮음

→ 또한, 샘플링 개수(support)도 비율
의 차이가 큼

선형보간법으로 온도 결측치에서
NG 값도 OK 범주에 있는 온도로
채웠기 때문에 기준이 모호해짐.

확실하지 않은 데이터를 이용해
모델의 성능이 떨어졌다고 판단함.

양품의 개수 7454
불량품의 개수 2882

03 결과분석

OK	Precision	Recall	f1-score	Support
0	0.77	0.75	0.76	720
1	0.91	0.91	0.91	1864

OK	Precision	Recall	f1-score	Support
0	0.60	0.36	0.65	720
1	0.79	0.91	0.94	1864

OK	Precision	Recall	f1-score	Support
0	0.78	0.77	0.77	867
1	0.91	0.91	0.91	2234

결측치를 모두 제거하고
진행하기로 결정함

0 : 불량품 1 : 양품

양품을 판별하는 건 높은 데에 비해
불량품을 판별하는 건 낮음

모든 샘플링 개수(support)도 비율
의 차이가 큼

보간법으로 온도 결측치에서
NG 값도 OK 범주에 있는 온도로
채웠기 때문에 기준이 모호해짐.

따라서 모델의 예측이 떨어졌다고
판단함.

03 피드백 적용

In [1]:

```
Users/macbookair/Documents/Tuk_Folder/AIManu/
```

In [4]: `df1 = df.dropna()`
`print(df1)`

...

In [5]: `df1.shape`

Out[5]: (9483, 6)

In [6]: `df1.isnull().sum()`

Out[6]:

STD_DT	0
MIXA_PASTEUR_STATE	0
MIXB_PASTEUR_STATE	0
MIXA_PASTEUR_TEMP	0
MIXB_PASTEUR_TEMP	0
INSP	0

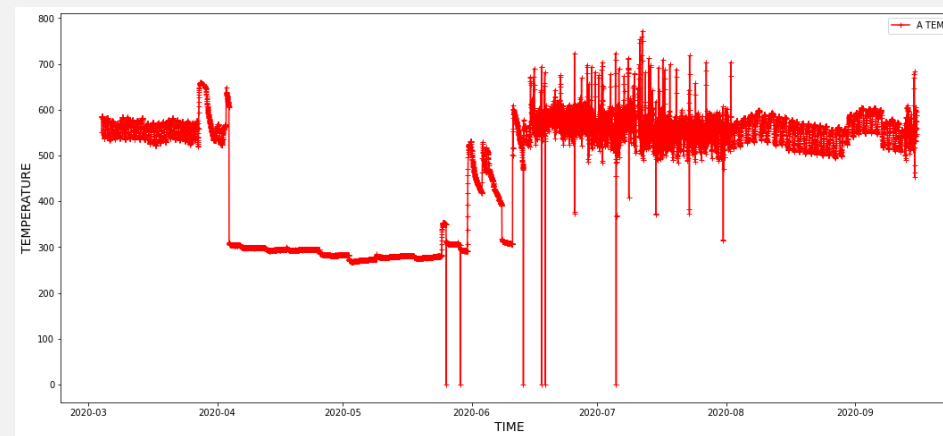
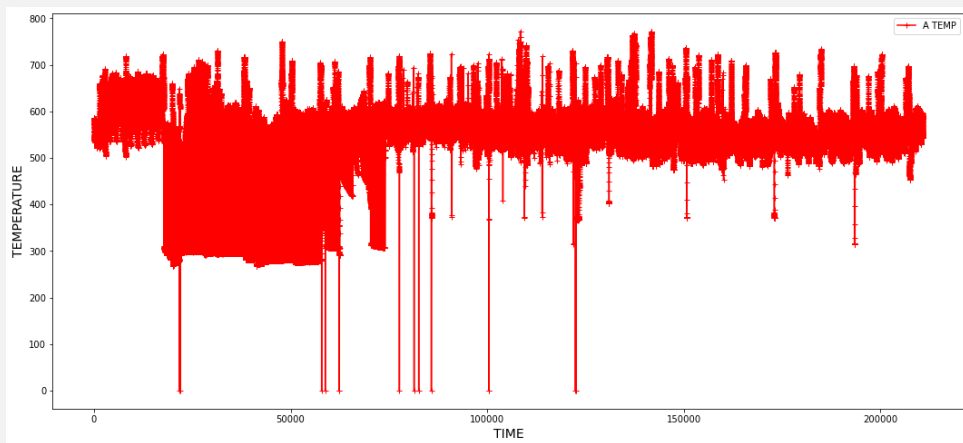
dtype: int64

1. 결측치를 모두 제거함

2. 이전과 동일하게 날짜표기법 수정 및
30분 단위시간으로 전처리 진행함

상태	개수
OK	7968
NG	1415

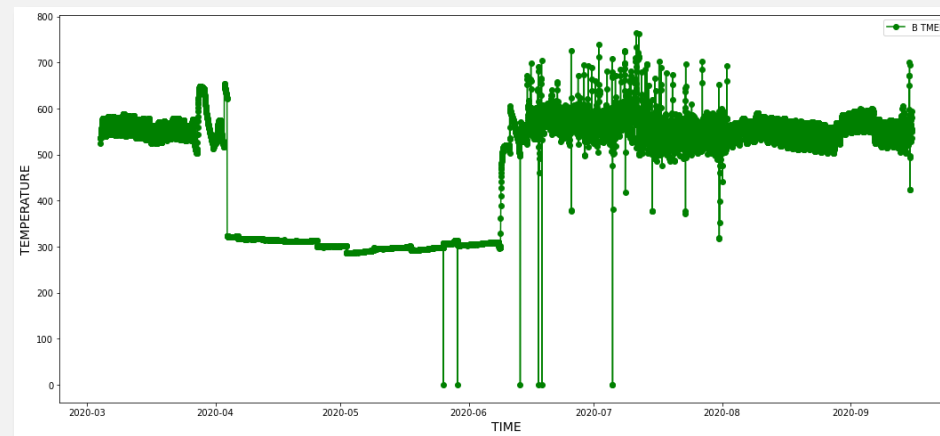
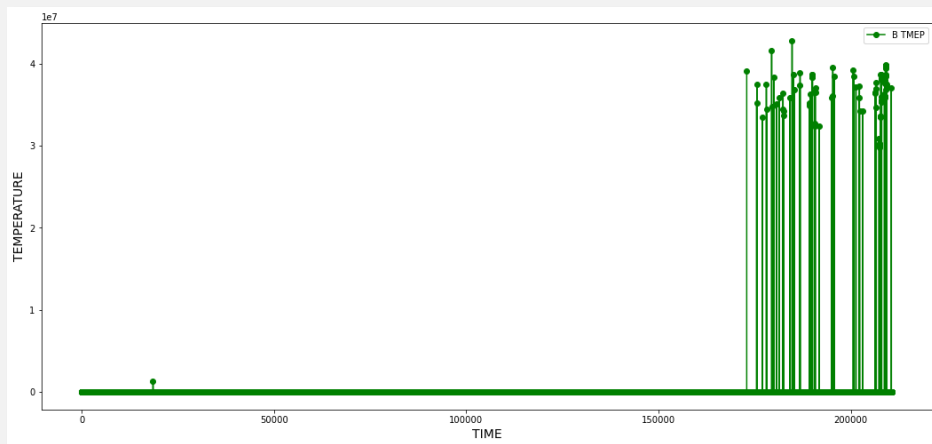
03 EDA



온도 A의 데이터 전처리 전후 과정

결측치를 제거한뒤 온도변화가 어떻게 되는지 시각화한 모습이다.

03 EDA



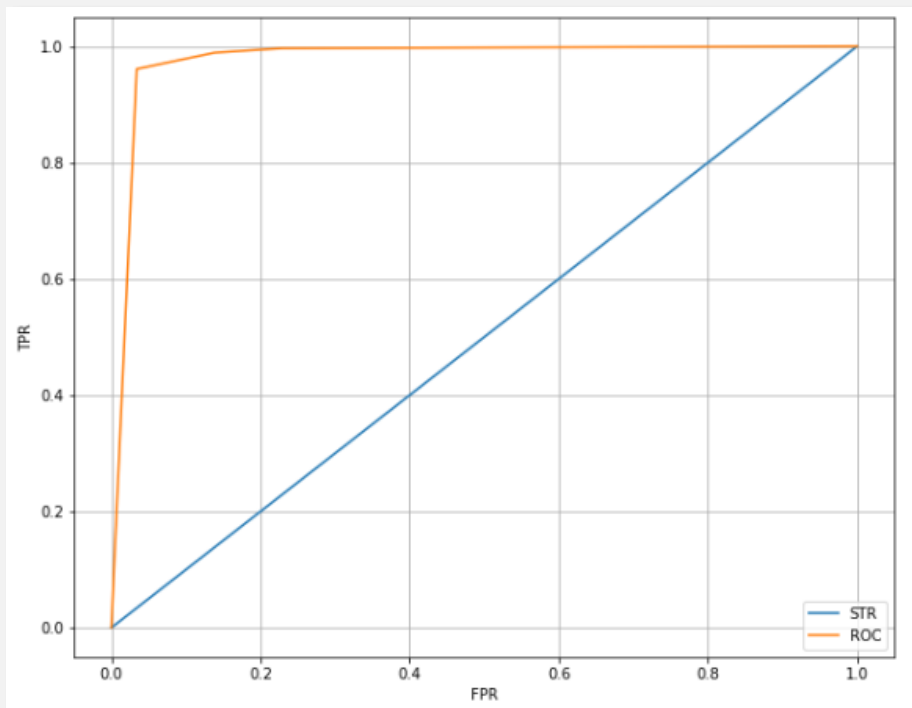
온도 B의 데이터 전처리 전후 과정

결측치를 제거한뒤 온도변화가 어떻게 되는지 시각화한 모습이다.

03 피드백 적용

모델	공통사항
KNN	1. Test size = 0.25 로 고정 2. Stratify = df['INSP'] 로, 원본 데이터 비율 유지하면서 Train_test_split 진행 3. 결측치 모두 제거
DNN	
Random Forest	

03 KNN



k	실행 시 정확도
1	0.97231097
2	0.97228145
3	0.97124601
4	0.97334755
5	0.97228145
6	0.97228145
7	0.97228145
8	0.96908316
9	0.97334755
10	0.97547974

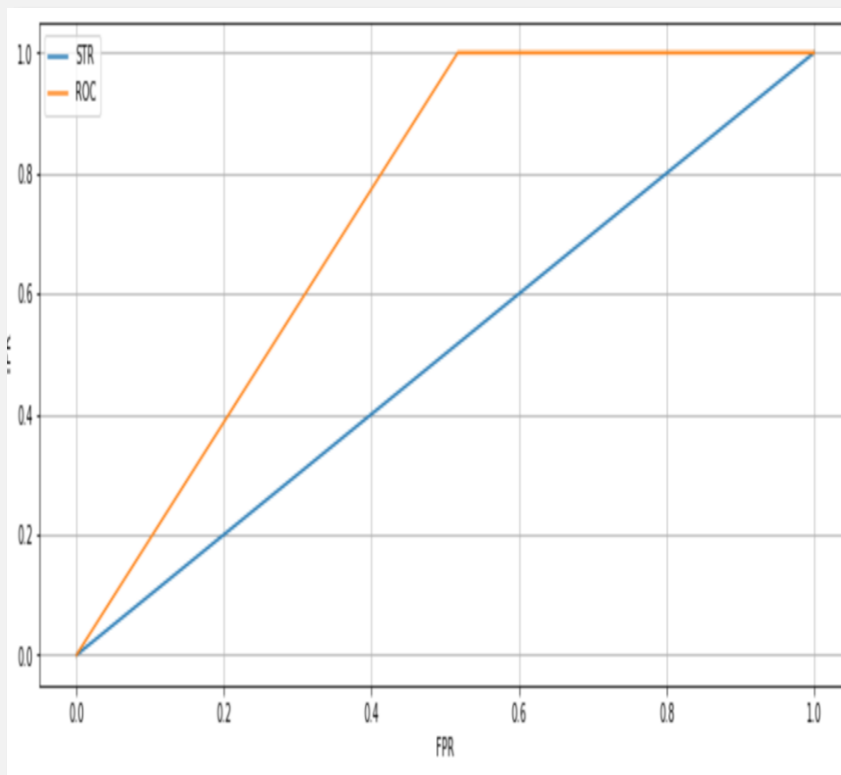
- 선형보간법 방법을 적용했을 때와 비교했을 때, 향상된 정확도와 성능을 확인할 수 있음
- 10 겹 교차검증을 수행했을 때의 정확도와 ROC 커브의 모습

03 KNN

	Precision	Recall	f1-score	Support
0	0.89	0.90	0.90	354
1	0.98	0.98	0.98	1952
정확도			0.97	2346
Macro avg	0.94	0.94	0.94	2346
Weighted avg	0.97	0.97	0.97	2346

Classification_report 를 이용해 모델 평가지표를 출력한 결과

03 DNN



k	실행 시 정확도
1	0.91349935
2	0.88335520
3	0.90956747
4	0.88713908
5	0.90944880
6	0.89501309
7	0.91601049
8	0.92125982
9	0.90944880
10	0.89370077

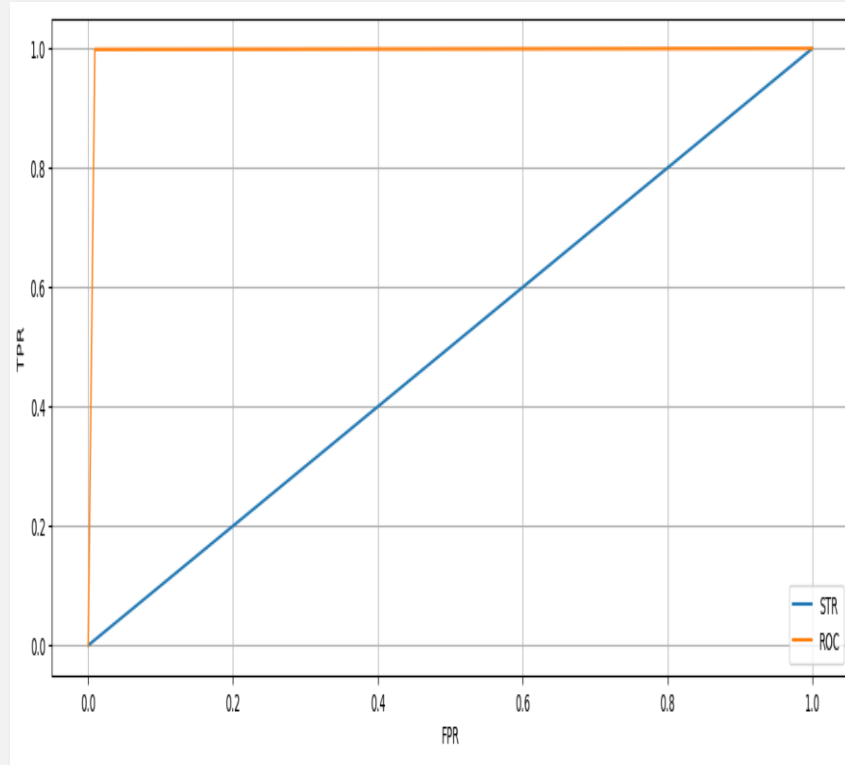
- 선형보간법 방법을 적용했을 때와 비교했을 때, 향상된 정확도와 성능을 확인할 수 있음
- 10 겹 교차검증을 수행했을 때의 정확도와 ROC 커브의 모습

03 DNN

	Precision	Recall	f1-score	Support
0	0.99	0.48	0.65	354
1	0.92	0.99	0.96	1992
정확도			0.92	2346
Macro avg	0.96	0.74	0.80	2346
Weighted avg	0.93	0.92	0.91	2346

Classification_report 를 이용해 모델 평가지표를 출력한 결과

03 Random Forest



k	실행 시 정확도
1	0.99680511
2	0.99787007
3	0.99
4	0.99
5	0.99893390
6	0.99573560
7	0.99786780
8	0.99893390
9	0.99
10	0.99680170

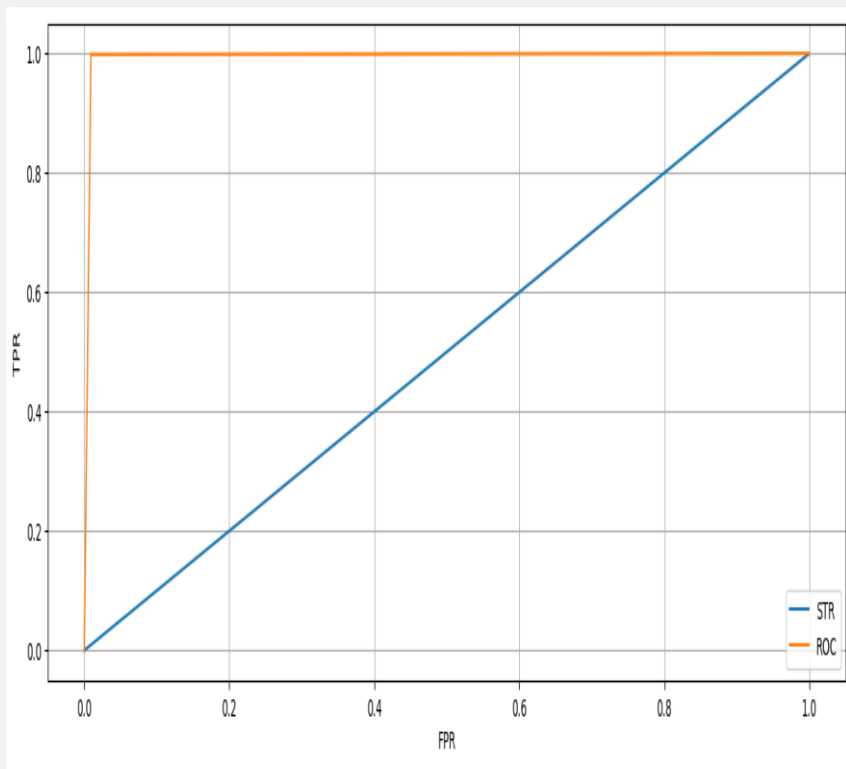
- 선형보간법 방법을 적용했을 때와 비교했을 때, 향상된 정확도와 성능을 확인할 수 있음
- 10 겹 교차검증을 수행했을 때의 정확도와 ROC 커브의 모습

03 Random Forest

	Precision	Recall	f1-score	Support
0	0.99	0.99	0.99	354
1	0.99	0.99	0.99	1992
정확도			0.99	2346
Macro avg	0.99	0.99	0.99	2346
Weighted avg	0.99	0.99	0.99	2346

Classification_report 를 이용해 모델 평가지표를 출력한 결과

03 모델 결정



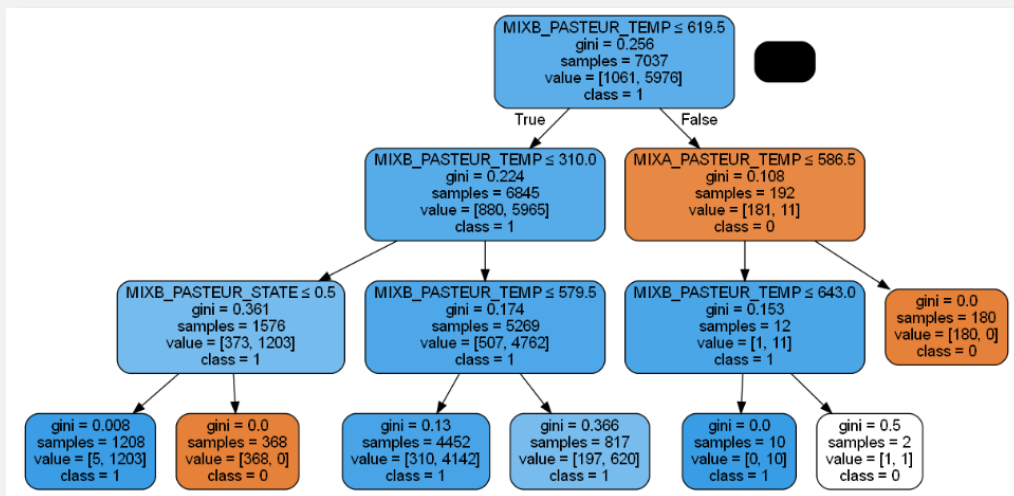
Random Forest

평균 정확도 : 0.998

- 모든 모델 중 가장 높은 정확도를 가지고 있으며 Roc 커브와 f1 score를 고려했을 때에도 성능이 가장 좋아 해당 모델로 선정함.

03 모델 해석

해석



다음은 랜덤 포레스트의 분류 모습이다.

1. MIXB_PASTEUR_TEMP가 619.5 초과하고 MIXA_PASTEUR_TEMP 586.5 이하인 경우 양품 분류
2. MIXB_PASTEUR_TEMP 619.5를 초과하고 MIXA_PASTEUR_TEMP 가 586.5 초과하는 경우 불량으로 분류
3. MIXB_PASTEUR_TEMP 가 이하에서도 양품으로 분류하지만 조건에 따라서 달라지므로 제대로 알 수 없음

03 모델 해석

해석

1. MIXB_PASTEUR_TEMP가 619.5 초과하고
MIXA_PASTEUR_TEMP 586.5 이하인 경우
양품 분류
2. MIXB_PASTEUR_TEMP 619.5를 초과하고
MIXA_PASTEUR_TEMP 가 586.5 초과하는 경우
불량으로 분류
3. MIXB_PASTEUR_TEMP 가 이하에서도 양품으로 분류하지만 조건에 따라서 달라지므로 제대로 알 수 없음

04 기대효과

모델 활용방안

1. 실제 설비시 **CCP 이행기준도 고려**해 온도 설정해야함
2. 다른 제품에 적용할 경우, 해당 **제품의 데이터로 분석 모델을 다시 학습**시켜야 됨
3. 살균공정에서 측정되는 **다른 변수들도 포함**해 학습시킨다면 **더 정확한 모델 생성**할 수도 있음
4. **제조현장마다 환경이 다르기에** 무조건 해당 모델이 맞을 수는 없음

기대효과

1. 데이터에 기반한 온도 조절 시스템을 가동하기 때문에 **장기적인 인건비 절약**과 **제품 품질 향상**을 기대할 수 있다.