

업무프로세스 개선을 위한 관리자의 의사결정 프로그램

현대중공업 DT 인력양성 프로그램

팀 쓸만한 놈들

IT 융합전공
전기전자공학전공

김민혁
윤형근

기계자동차공학전공
IT 융합전공

권경호
이세현

산업경영공학전공

김창영

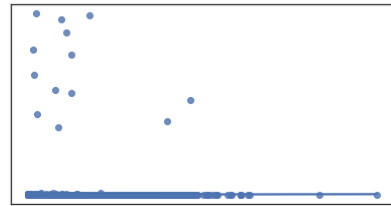
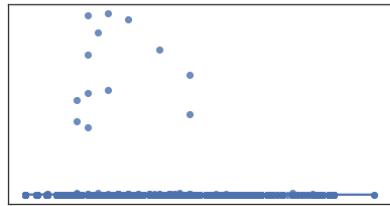
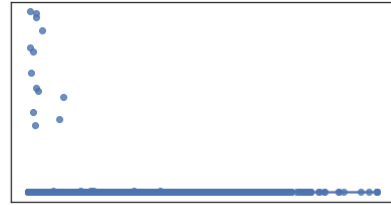
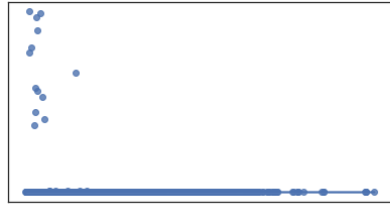
01^{- 01} CNC 작업

1. 데이터 시각화
2. 분산분석
3. 사후분석
4. 변수선정
5. 이상치 탐지
6. 모델설계
7. 결과분석

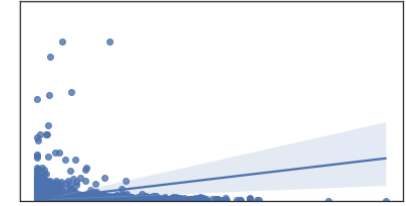
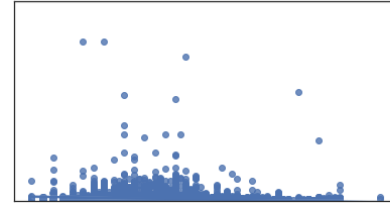
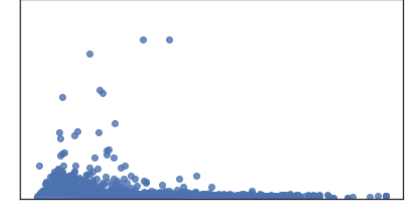
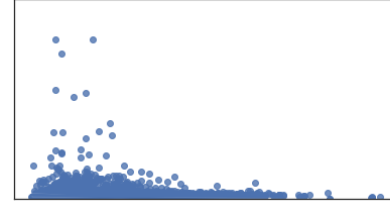


데이터 시각화(수치형)

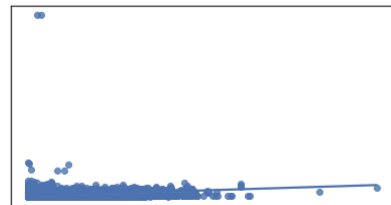
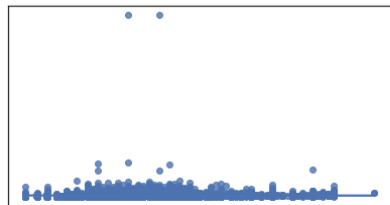
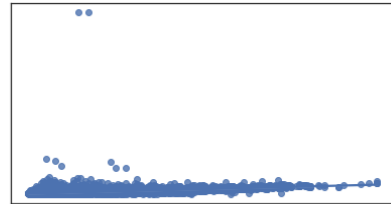
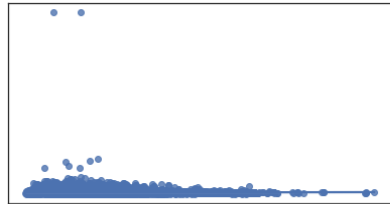
■ Y1



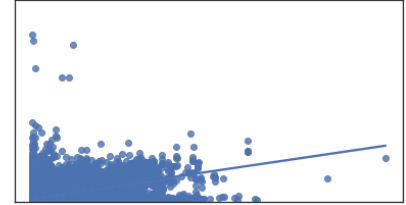
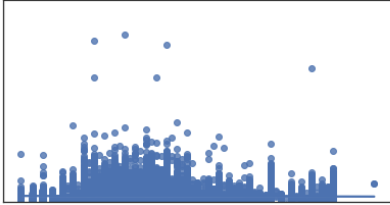
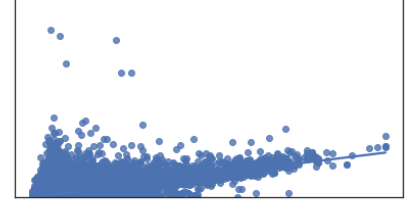
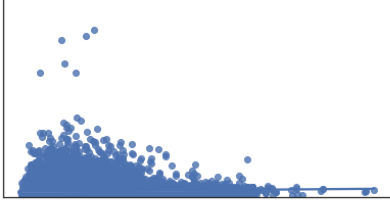
확대
→



■ Y2

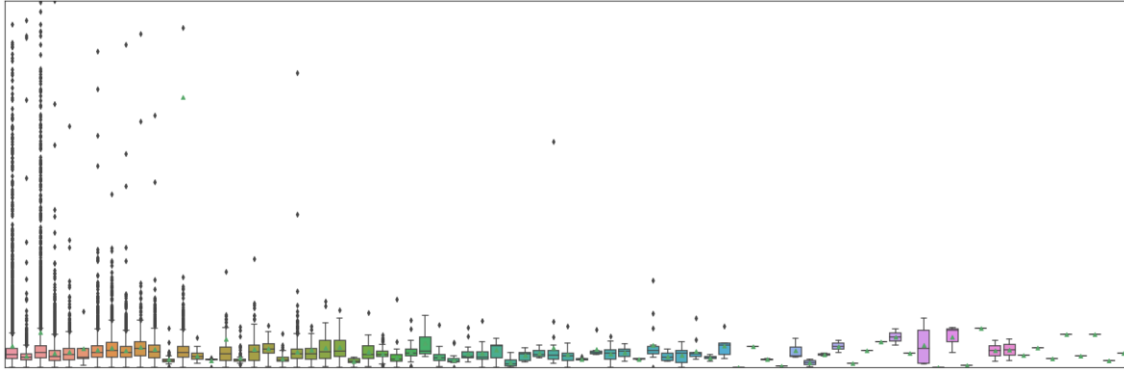


확대
→

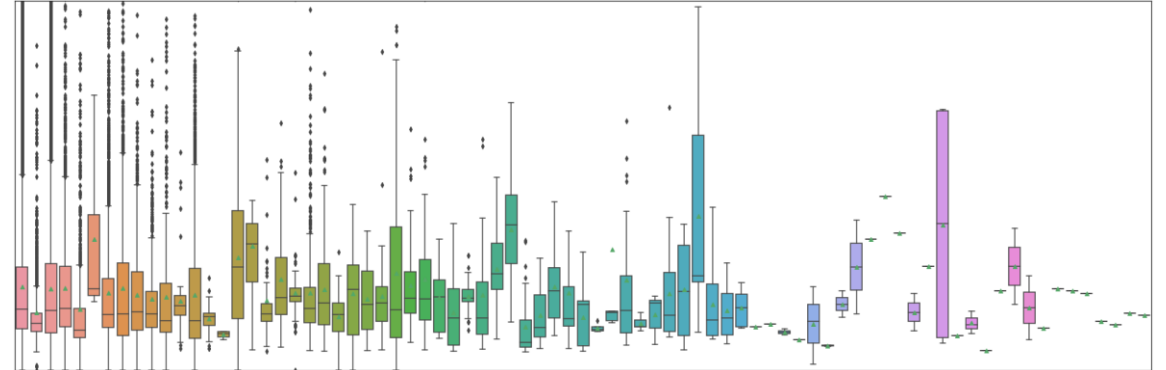




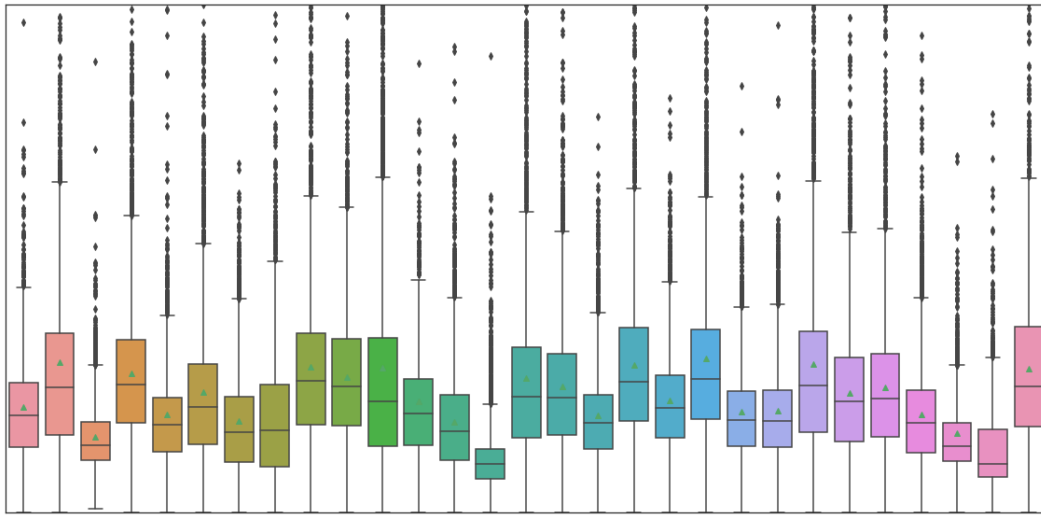
데이터 시각화(범주형)



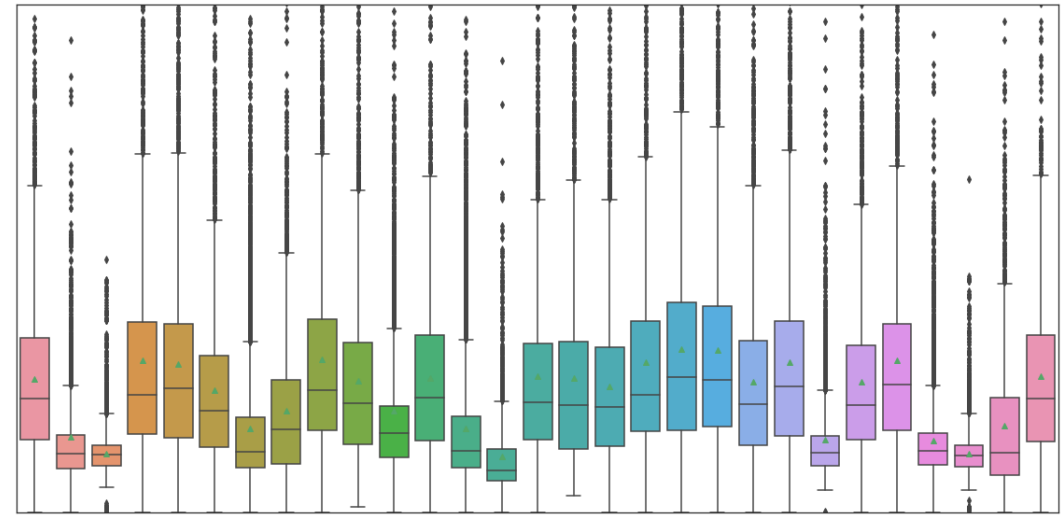
[그림1-1] X5~Y1



[그림1-2] X5~Y2



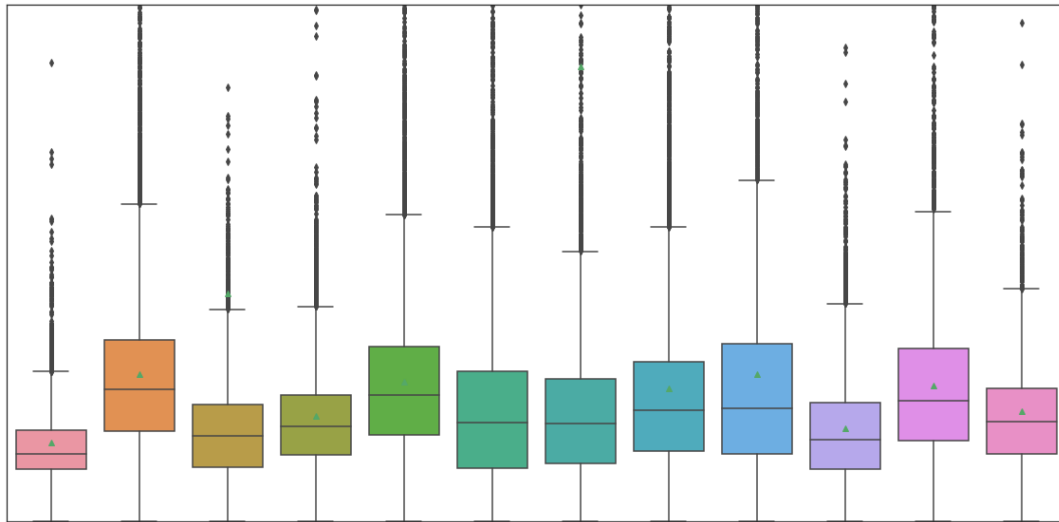
[그림1-3] X8~Y1



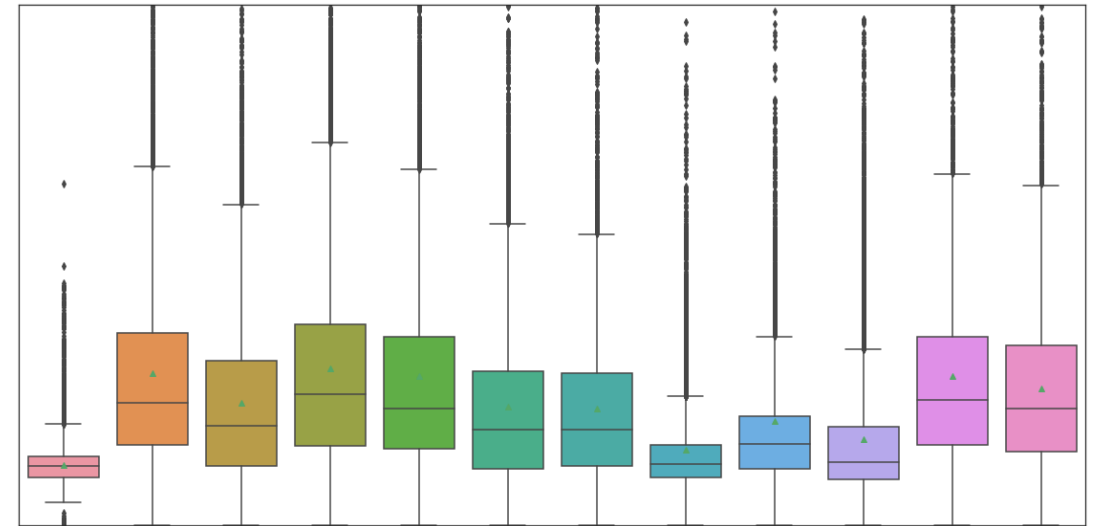
[그림1-4] X8~Y2



데이터 시각화(범주형)



[그림1-5] X7~Y1



[그림1-6] X7~Y1



분산분석

- 예측변수와 범주형 데이터들의 분산분석 결과(P-value)
 - 유의수준 0.05

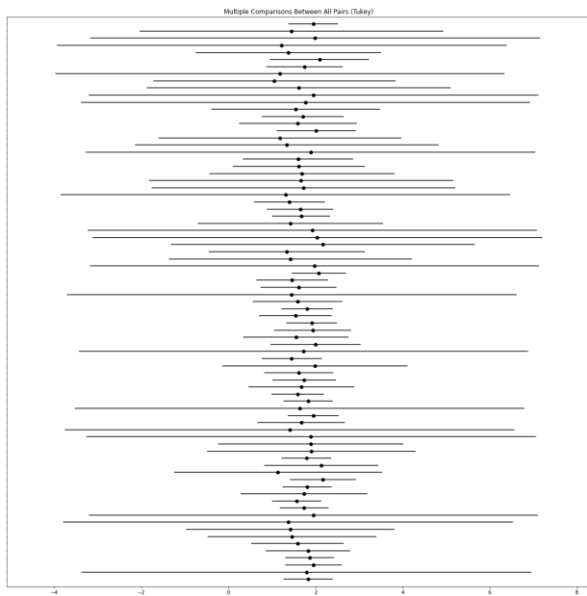
[표1] 분산분석 결과

	X5	X7	X8
Y1	1.0	3.6686e-11	1.7218e-55
Y2	7.3222e-30	0.0	0.0

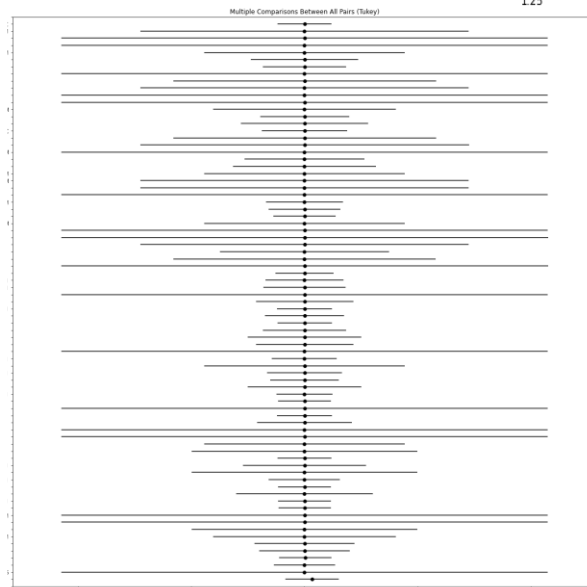


사후분석(Post-hoc)

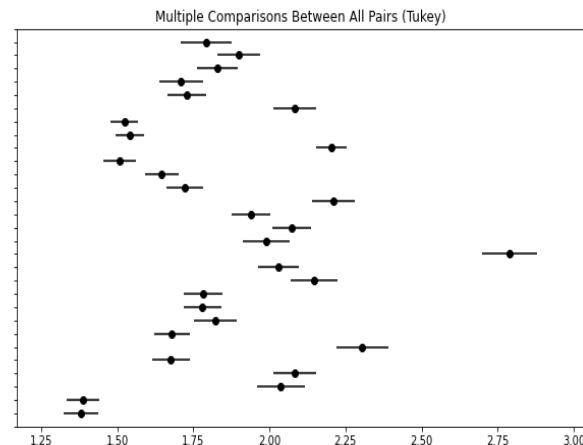
- 범주형 데이터 처리
 - 유의한 변수들 Target encoding(평균으로 변환)
 - 유의하지 않은 변수들 one-hot encoding



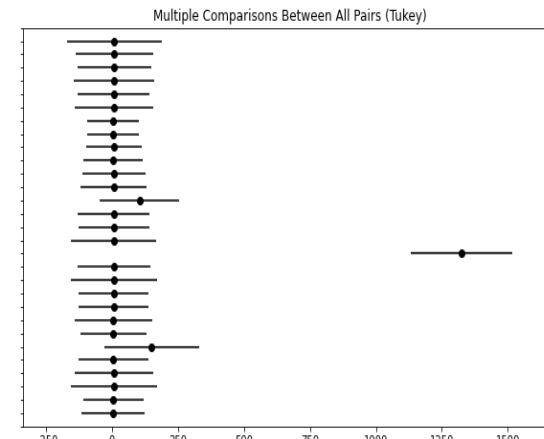
[그림2-3] X5 - Y2



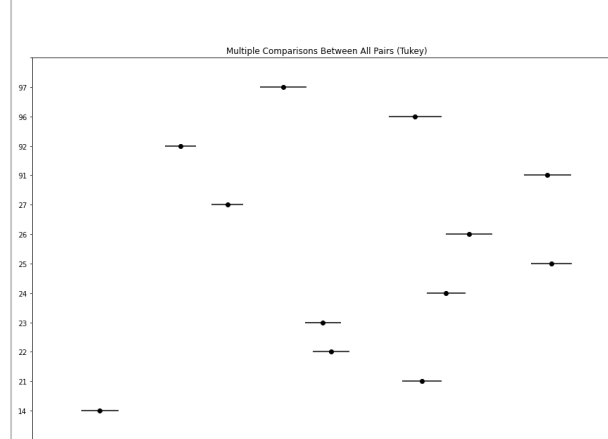
[그림2-4] X5 - Y1



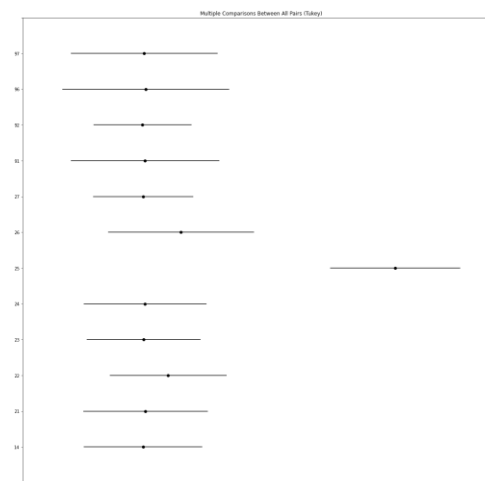
[그림2-1] X8 - Y2



[그림2-2] X8 - Y1



[그림2-5] X7 - Y2



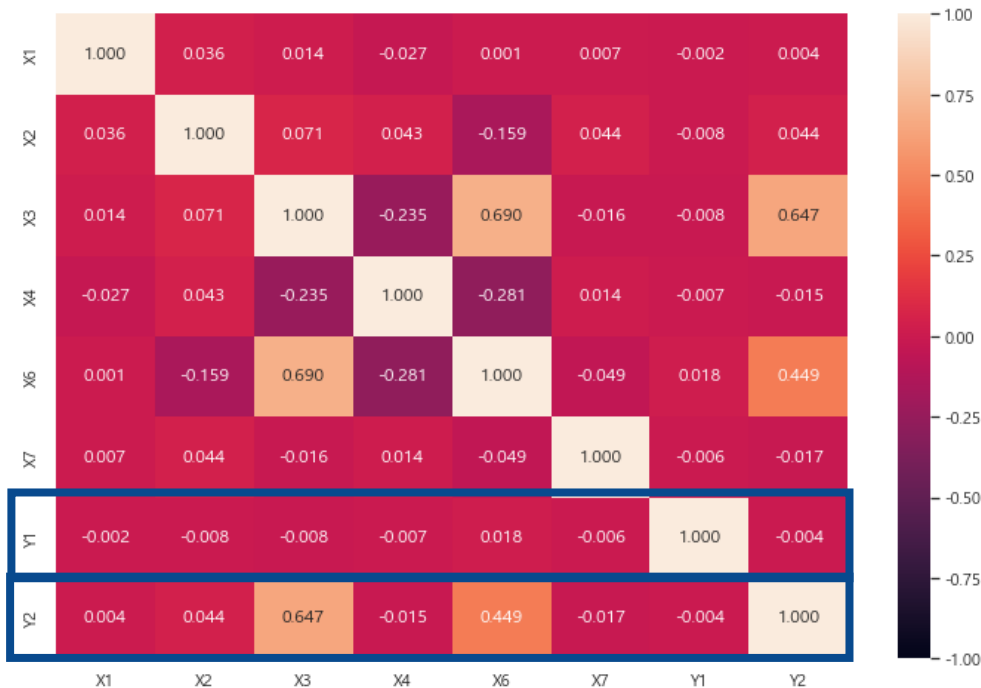
[그림2-6] X7 - Y1



상관분석

■ 상관분석

[표2-1] 범주형 데이터 전처리 전 상관분석



[표2-2] 범주형 데이터 전처리 후 상관분석



- Y1 – X2, X8_mt, X7_mt, X3, X5_mt, X4
- Y2 – X3, X6, X8_ct, X7_ct, X5_mt, X4



변수선택

- Variance Inflation Factors(VIF)

VIF \geq 10 다중공선성 유발 가능

- 마킹시간

	Feature	VIF_Factor
0	X2	3.920559
1	X8_mt	83.797542
2	X7_mt	105.158866
3	X3	3.289605
4	X5_mt	33.914724
5	X4	12.841500

	Feature	VIF_Factor
0	X2	3.877579
1	X8_mt	19.032786
2	X3	3.159293
3	X5_mt	20.049691

	Feature	VIF_Factor
0	X2	3.696856
1	X8_mt	5.319574
2	X3	3.035482

	Feature	VIF_Factor
0	X2	3.782448
1	X5_mt	5.603794
2	X3	3.113392

- 절단시간

	Feature	VIF_Factor
0	X3	7.111851
1	X6	3.410723
2	X8_ct	141.342487
3	X7_ct	155.411552
4	X5_ct	32.136144
5	X4	15.726553

	Feature	VIF_Factor
0	X3	6.433302
1	X6	2.786674
2	X8_ct	14.155720
3	X5_ct	29.696292
4	X4	15.707818

	Feature	VIF_Factor
0	X3	6.424566
1	X6	2.691025
2	X8_ct	14.139992
3	X5_ct	12.118548

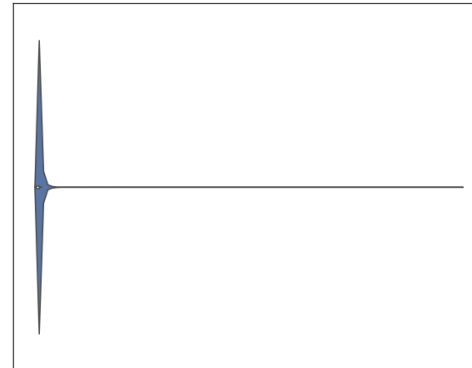
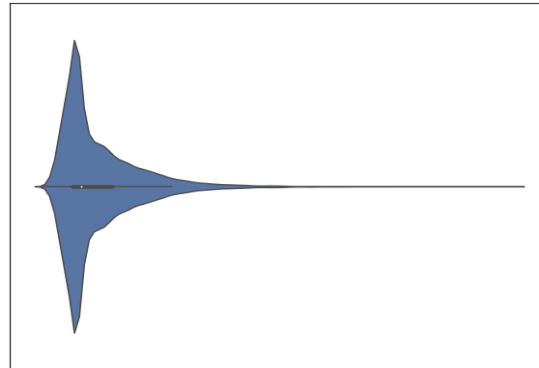
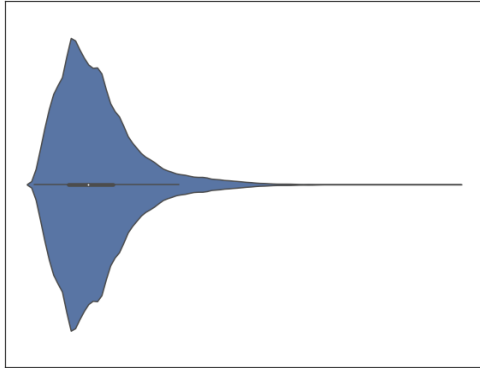
	Feature	VIF_Factor
0	X3	6.385731
1	X6	2.760959
2	X8_ct	9.546682
3	X4	6.410092



이상치탐지

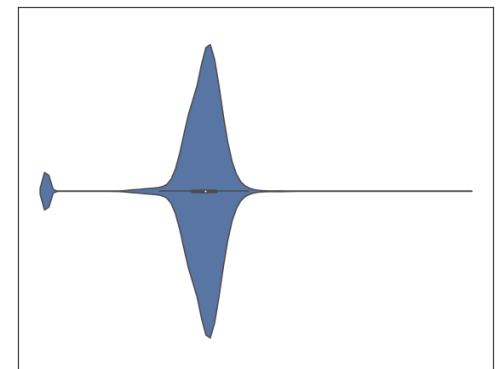
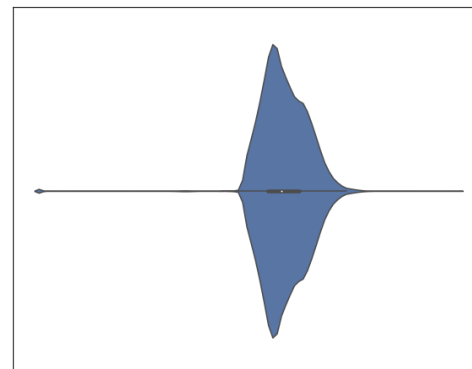
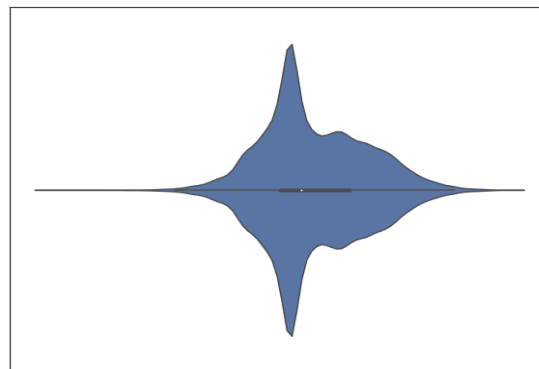
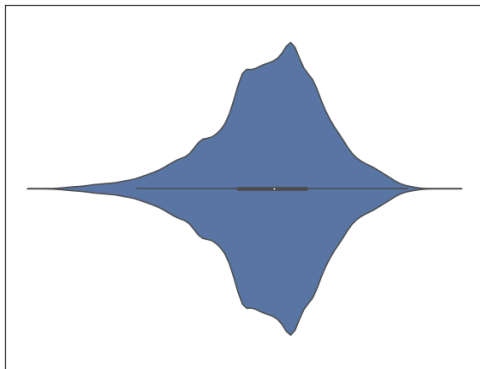
로그 스케일링

변환 전



[그림3-1] 변환 전 분포도

변환 후



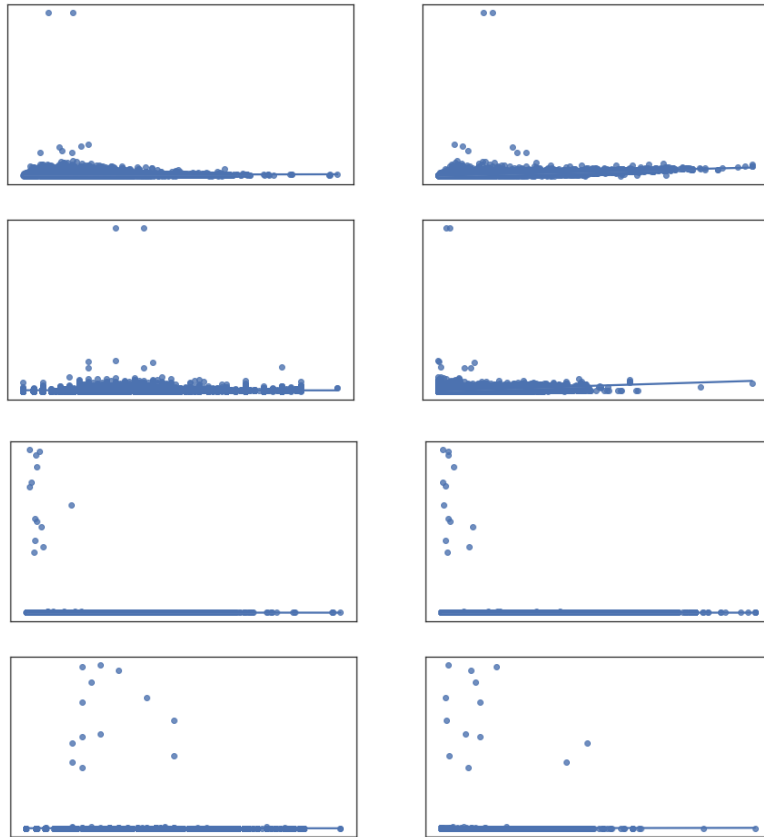
[그림3-2] 변환 후 분포도



이상치 탐지

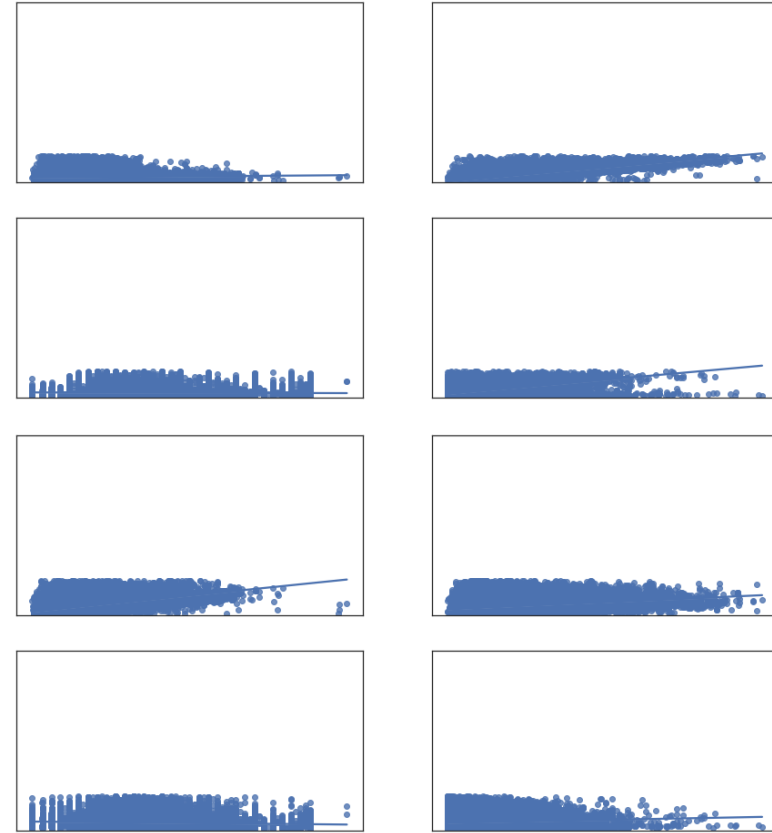
■ 4분위수 기반 이상치 탐지

■ 이상치 제거 전



[그림4-1] 이상치 제거 전 산포도

■ 이상치 제거 후

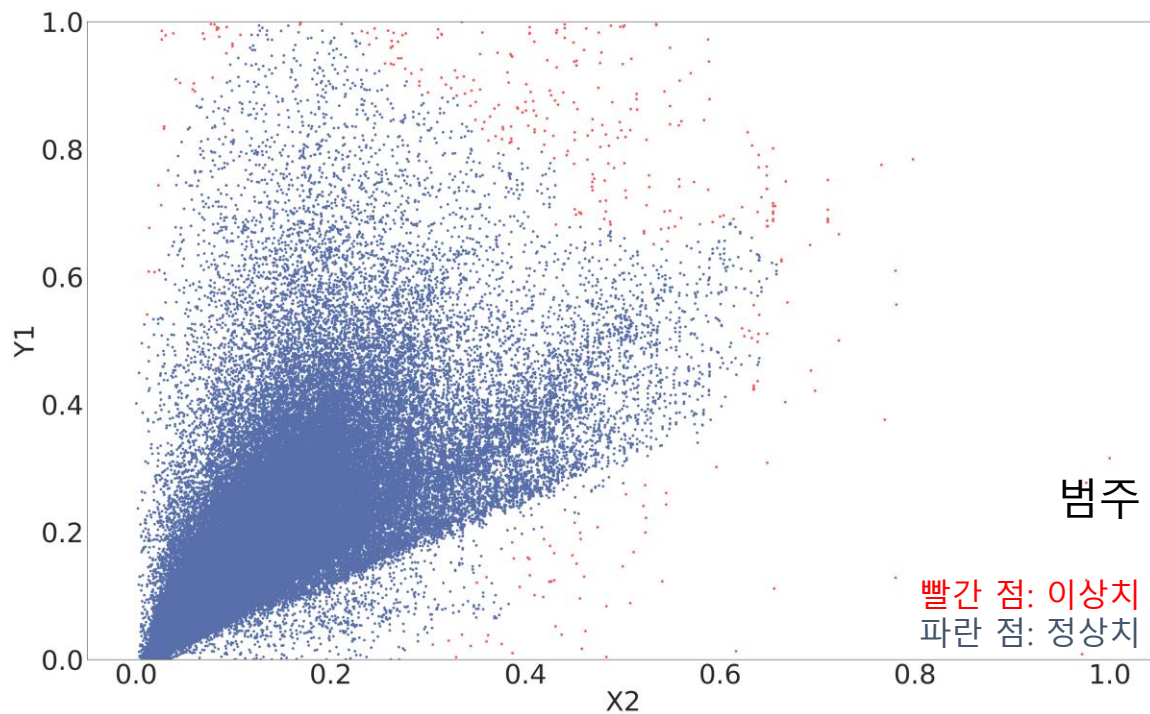


[그림4-2] 이상치 제거 후 산포도

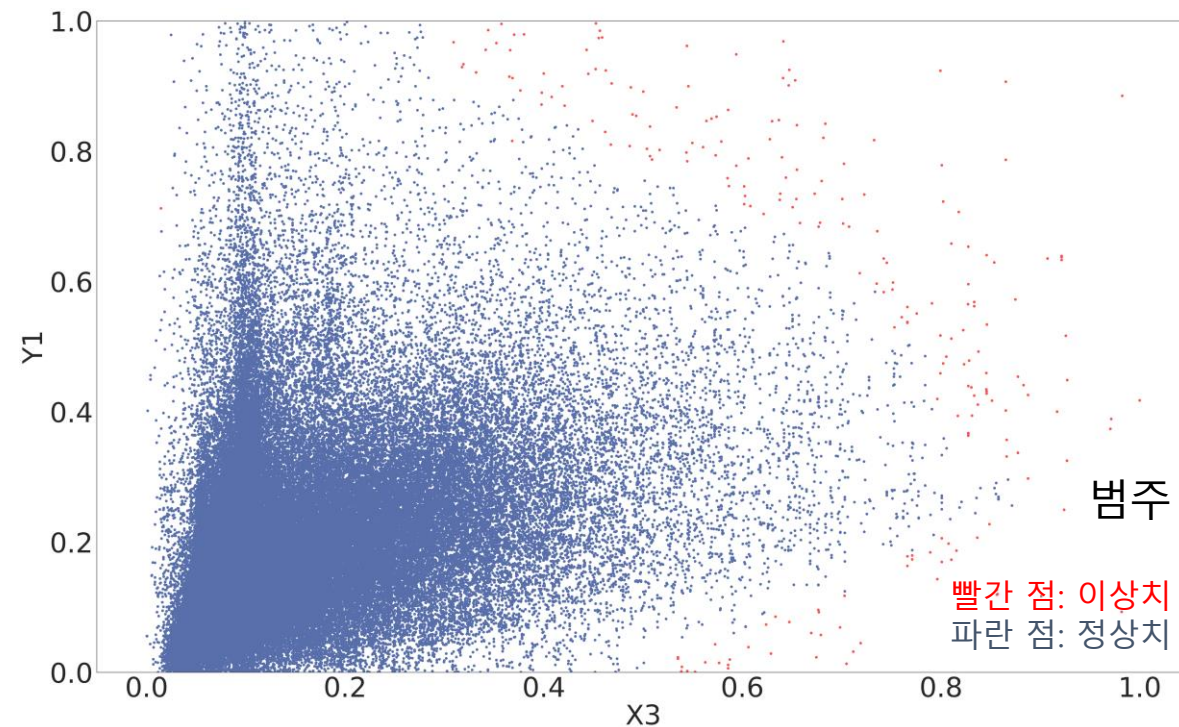


이상치 탐지

- DBSCAN 기반 이상치 탐지(Y1)



[그림5-2] X2-Y1

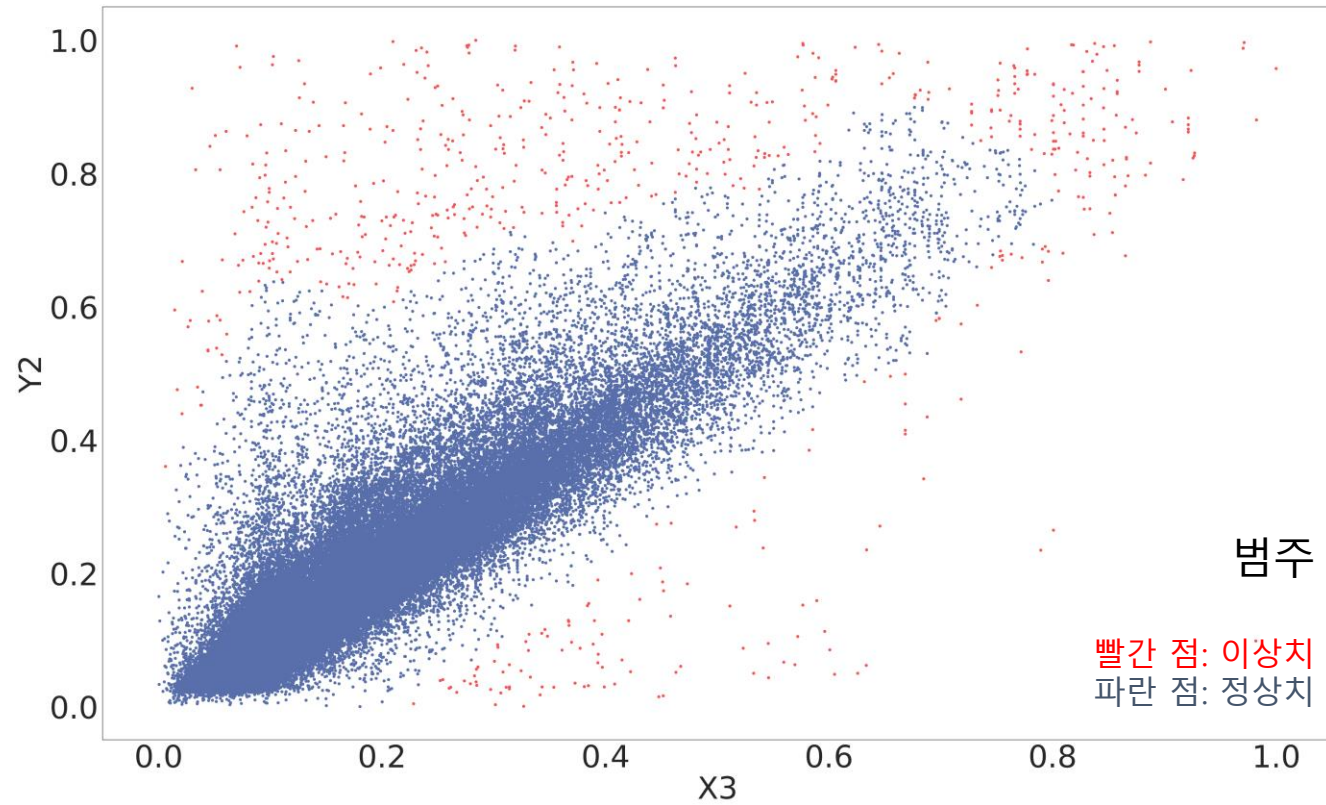


[그림5-3] X3-Y1



이상치 탐지

- DBSCAN 기반 이상치 탐지(Y2)



[그림5-1] X3-Y2



모델 설계

■ Y1

■ 변수선택

X2, X3, X5(one-hot-encoding), X8(Target-encoding)

■ 모델

z-score 정규화

Xgboost, Lightgbm, Random Forest -> Hard Voting

교차검증: 10-fold 진행

■ Y2

■ 변수선택

X3, X6, X4, X5(Target-encoding) , X8(Target-encoding)

■ 모델

z-score 정규화

Lightgbm -> Bagging

교차검증: 10-fold

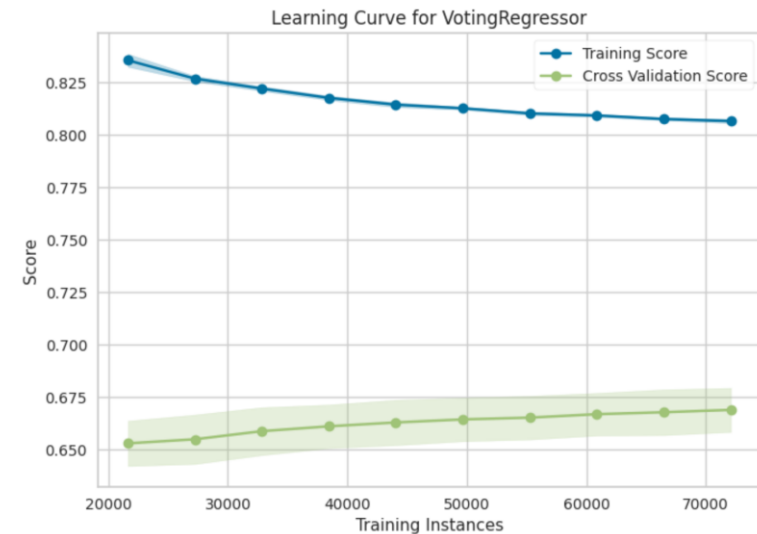
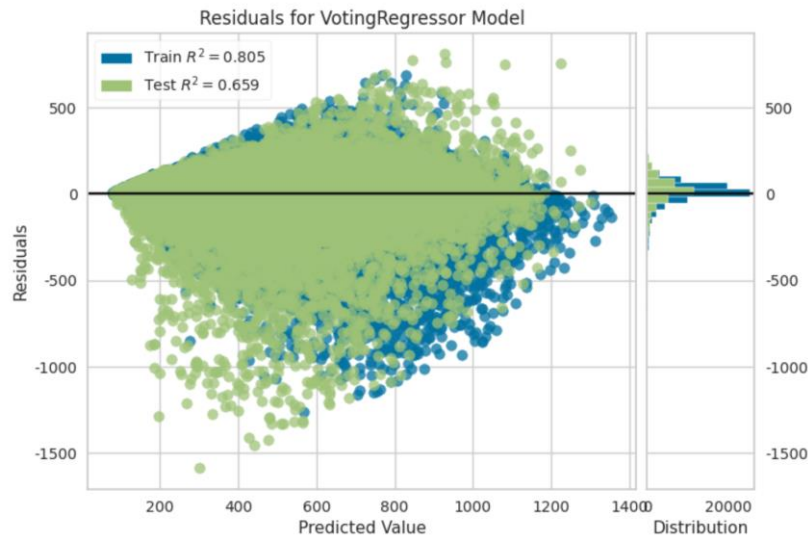
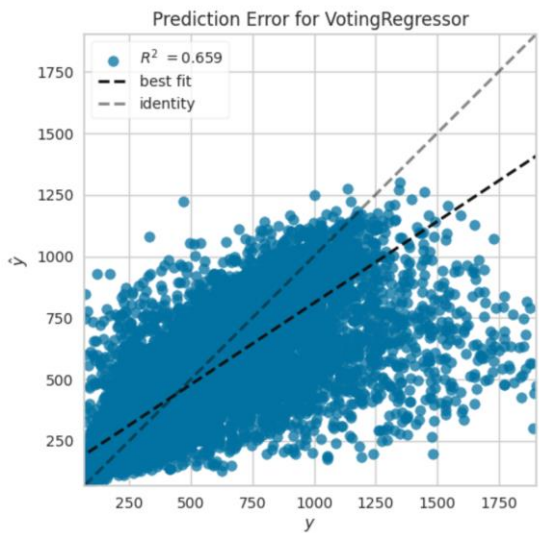


결과 분석

■ Y1

[표3-1] Y1 모델 결과

	MAE	RMSE	R2
결과값	79.9579	140.0765	0.6690



[그림6-1] Y1 모델 결과 시각화

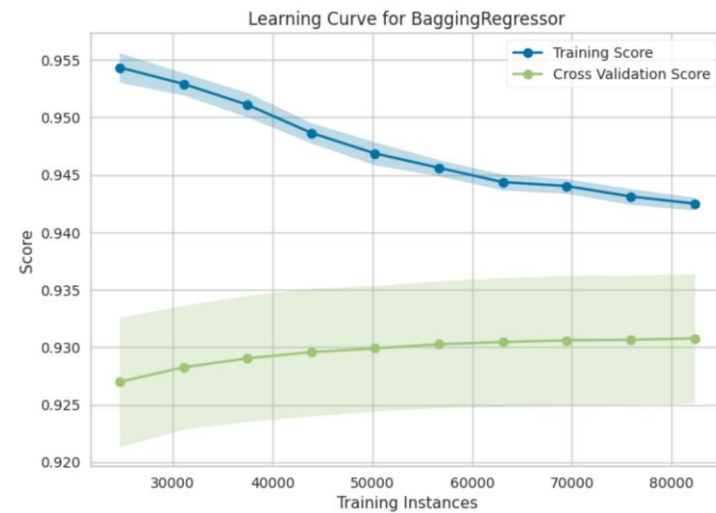
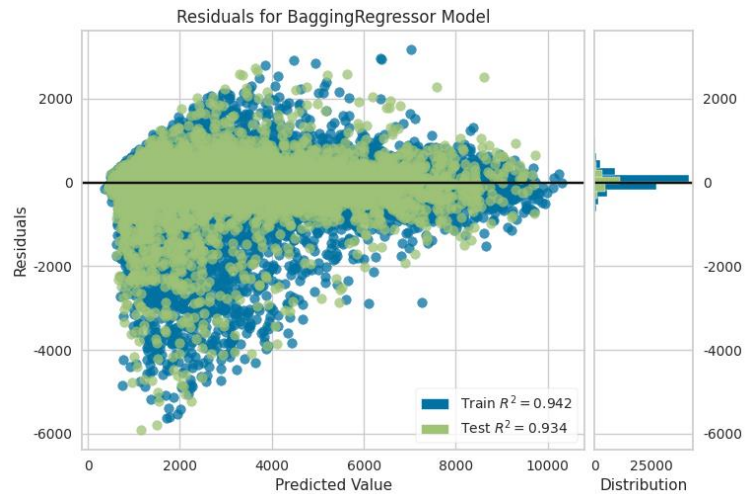


결과 분석

■ Y2

[표3-2] Y2 모델 결과

	MAE	RMSE	R2
결과값	179.7838	385.7666	0.9307



[그림6-2] Y2 모델 결과 시각화

01^{- 02} CNC 작업할당

1. 문제 및 성과척도 정의
2. Greedy 알고리즘
3. 정수계획법
4. 결과비교



문제 및 성과척도 정의

- 성과척도

- 1. 총 완료시간

- 2. 기계 가동율

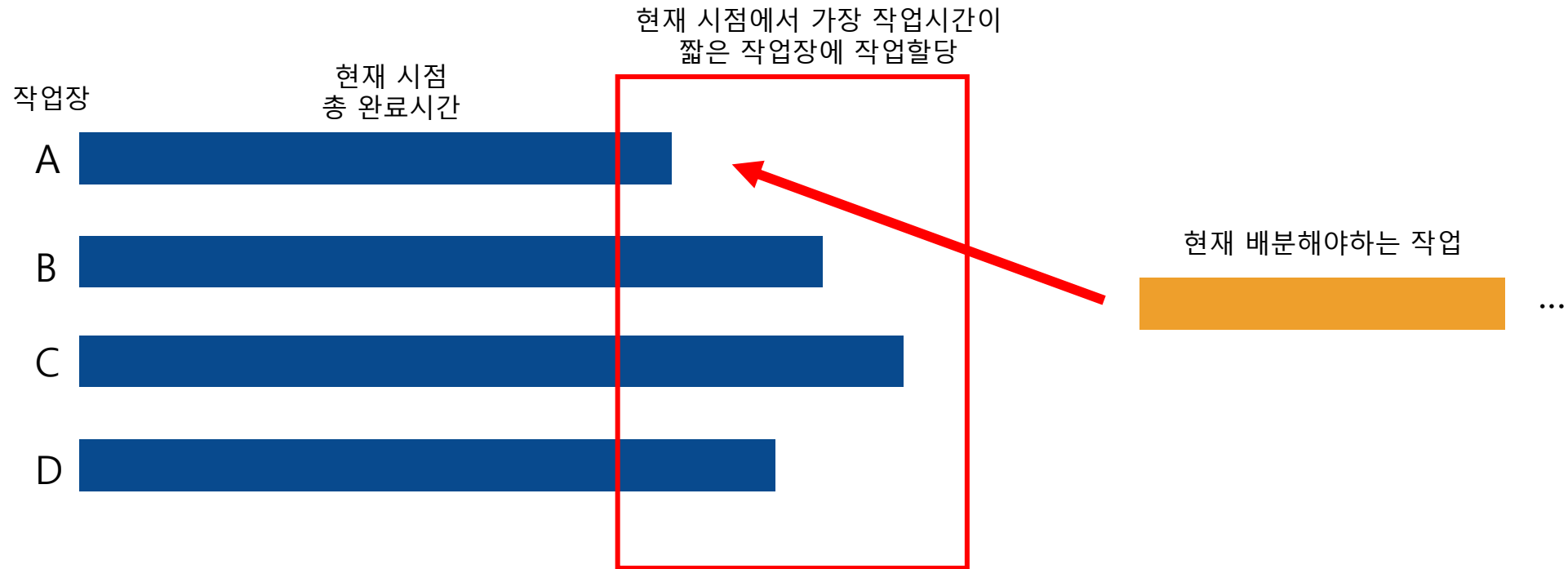
- 총 유휴시간 = (기계의 수) x (총 완료시간) 총처리시간
 - 가동률 = (총 시간 - 총 유휴시간) / 총 시간

- 문제 설계를 위한 가정

- 작업은 작업장의 설비와 관계없이 배분이 가능하다.
 - 작업장의 일일 작업은 6시간을 넘길 수 없다.
 - 작업중 휴식시간은 고려하지 않는다.



Greedy 알고리즘



[그림7] 그리디 알고리즘 방법론



정수 계획법(IP)

기존 IP모형의 목적함수

목적함수:

$$\min \max \sum_{j=1}^l Vx_{ij} \text{ (행 벡터 } V = [\text{절단시간(작업)}])$$



임의의 변수 t 도입
Equivalence

변경 후 IP모형의 목적함수

목적함수:

$$\min t$$

작업장 (총 k개)

(절단시간)작업 (총 l개)

	0	1	2	3	4	5	6	7	8	9	...	90	91	92	93	94	95	96	97	98	99
X7																					
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
91	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
92	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
96	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
97	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

[그림8] 작업장-절단시간(작업) Matrix

작업장에 해당 작업이 할당되면 1로 표기, 할당되지 않으면 0으로 표기



정수 계획법(IP)

IP모형

목적함수:

$$\min t$$

제약조건:

$$0 \leq x_{ij} \leq 1 \quad (\text{단 } x_{ij} \text{는 정수, } 1 \leq i \leq k(\text{작업장의 수}), 1 \leq j \leq l(\text{작업의 수}))$$

$$V = [v_1, v_2, v_3, v_4, \dots, v_j] \quad (\text{작업 시간의 행벡터})$$

$$E(V) \leq t$$

$$\sum_{j=1}^l v_j x_{1j} < t, \quad \sum_{j=1}^l v_j x_{2j} < t, \quad \sum_{j=1}^l v_j x_{3j} < t \quad \dots \quad \sum_{j=1}^l v_j x_{kj} < t$$

$$\sum_{i=1}^k x_{i1} = 1, \quad \sum_{i=1}^k x_{i2} = 1, \quad \sum_{i=1}^k x_{i3} = 1 \quad \dots \quad \sum_{i=1}^k x_{il} = 1$$

작업장 (총 k
개)



[그림9] IP 모형 목적함수

(절단시간)작업 (총 l개)

	0	1	2	3	4	5	6	7	8	9	...	90	91	92	93	94	95	96	97	98	99
X7																					
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
91	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
92	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
96	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
97	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

[그림8] 작업장-절단시간(작업) Matrix



결과비교

[표4] 작업할당 알고리즘 결과

	총 완료시간(s)	프로그램 구동시간
그리디 알고리즘	19459	1s
정수계획법(IP)	19340	60m (Timeout)

■ 결과 비교

- 성과척도: 정수계획법 > 그리디 알고리즘
- 각 절단시간별 편차가 크면 클수록 결과가 더 차이남

■ 정수계획법의 최적해 찾는 시간 보완

- 휴리스틱 방법론

가장 이상적인 최적해(모든 절단시간(작업)의 합/ 작업장의 수)에서 허용오차를 줄으로써 최적해 탐색

-> 시간 단축 효과

02 화물적재

1. 사전 계획
2. 데이터 전처리
3. 모델학습
4. 예측결과
5. 결과분석



사전계획

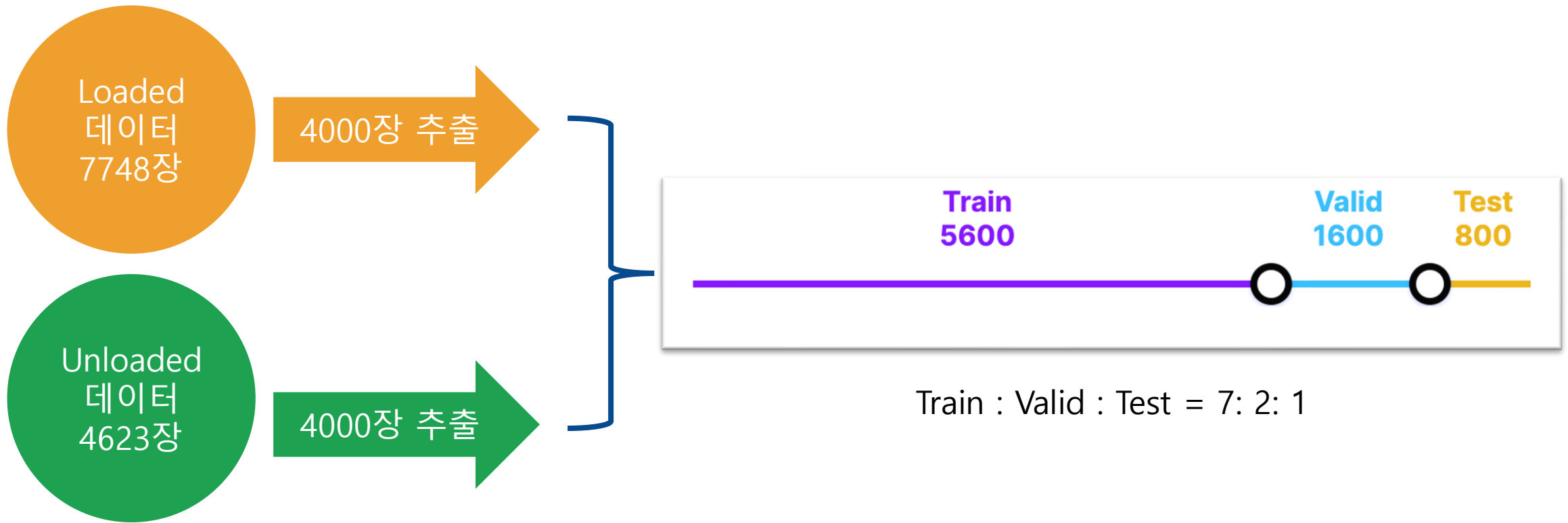
- Object Detection

- 데이터에서 화물 칸을 객체검출을 통해 찾아내고 화물 적재 유/무를 판단
- YOLO 모델 활용

- Image Classification

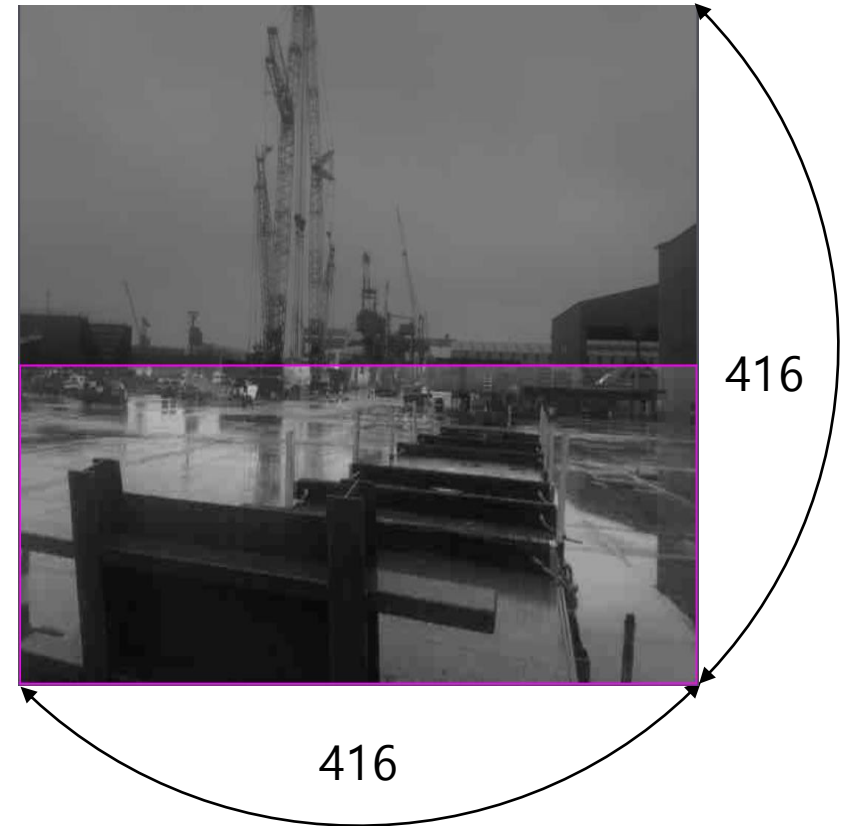
- 데이터에서 바로 화물 적재 유/무를 판단
- CNN 모델 활용

데이터 전처리



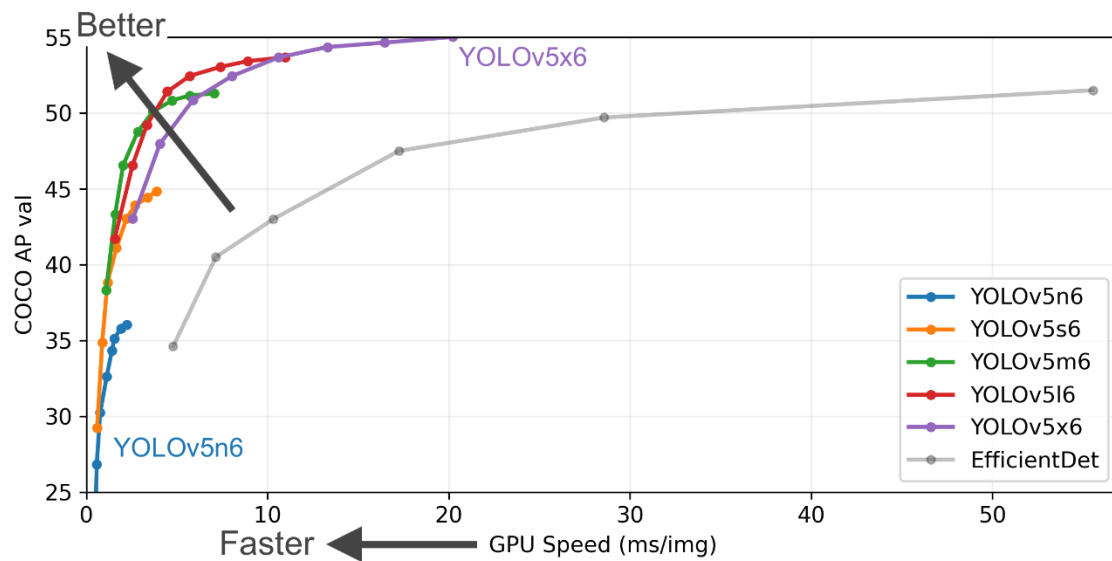


데이터 전처리



- Grayscale 적용
- Box 형태의 Layer 적용 Annotation 진행

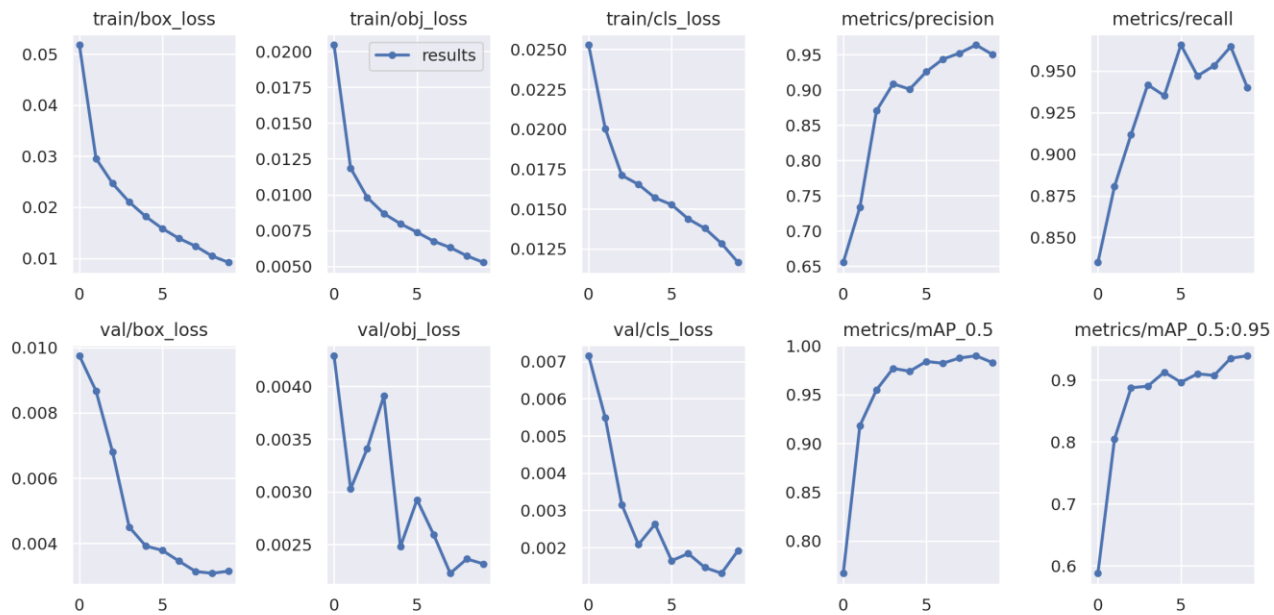
모델 학습





결과분석

YOLO_V5 모델 학습 결과



[그림11] YOLOv5 모델 결과



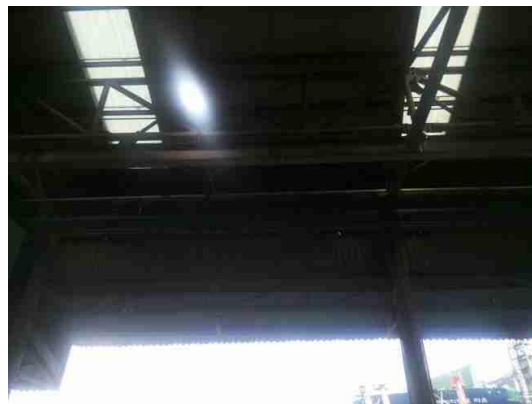
[그림12] Confusion Matrix

- 과적합의 문제 발생
- 신뢰도에 따라 한 이미지에서 아래와 같이 클래스 판단 오류 발생
- 최종 CSV 출력 과정에서 누락되거나, 중복으로 판단된 이미지 다수 존재



데이터 전처리

- 학습에 방해되는 데이터 제거

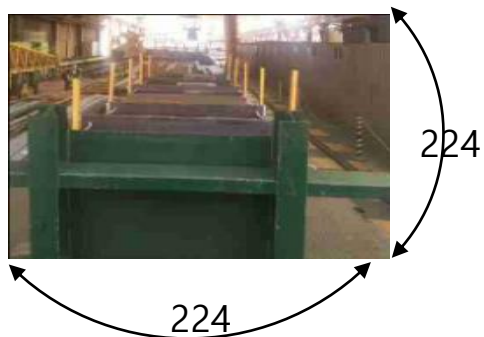




데이터 전처리

- 기존 훈련데이터의 개수
 - 화물이 적재되지 않은 데이터 7748장, 화물이 적재된 데이터 4623장
 - 총 12371장의 데이터
- 정리 이후 훈련데이터의 개수
 - 화물이 적재되지 않은 데이터 3128장, 화물이 적재된 데이터 1797장
 - 총 4919장의 데이터

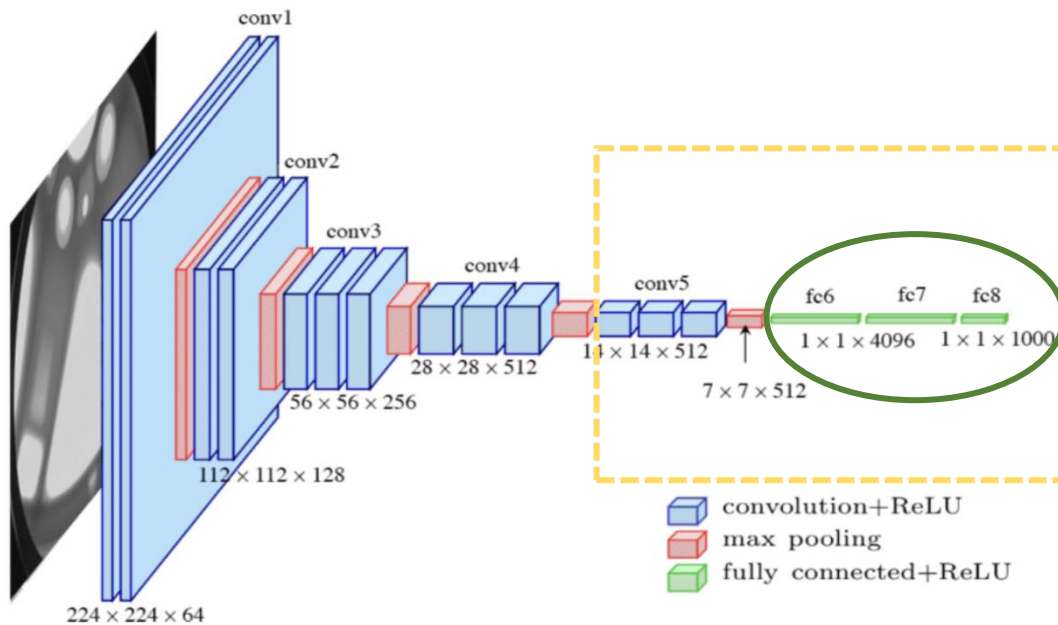
데이터 전처리



- 가로 이미지 크기의 15%, 세로 이미지 크기의 50% 추출
- 회전, 이동, 대칭, 밝기조절을 통해 데이터 증강

모델설계

■ VGG16



[그림13] 모델 도식화

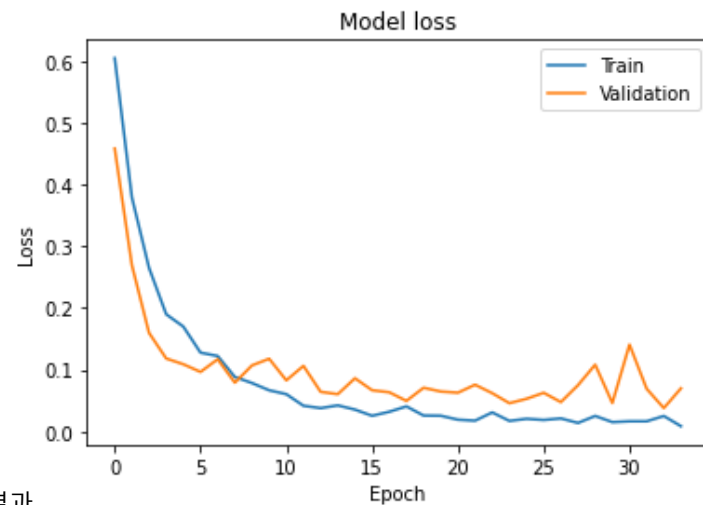
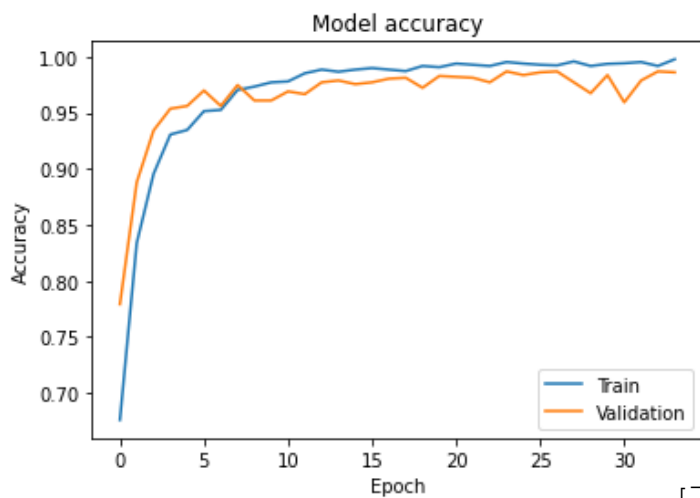
- 화물 적재 유무를 구분하기 위해 최상위 계층을 새롭게 정의
- Dense layer를 사용하며 sigmoid 함수를 사용하도록 변경
- 모델을 데이터와 잘 맞도록 조정하기 위하여 conv5 층부터 동결을 해제하고 학습



결과분석

■ CNN 기반 모델 학습 결과

- 검증 데이터셋에서의 정확도 98%



[그림14] CNN 모델 결과



결과분석

- 테스트 데이터셋에서 검증 데이터셋보다 정확도가 낮아진 이유
 - 과적합 문제
 - 학습데이터를 일괄적으로 같은 비율로 자르면서 필요한 정보가 많이 사라짐



[그림15] 데이터 전처리 사진

03 마무리

최적화
+
데이터 제공



시간 단축



업무
프로세스
개선



빠른
의사결정

A white flag with a green triangle and a yellow stripe, waving in the wind. The text "감사합니다" is overlaid in the center.

감사합니다