

---

***HD현대 AI Challenge***

---

**〈본선〉**

# Context

1

## INTRO

- 1) 문제 이해
- 2) 문제 제기

2

## ANALYSIS

- 1) 도메인 분석
- 2) 데이터 분석
- 3) 시계열 한계
- 4) 데이터 가공

3

## MODEL

- 1) 사용 모델
- 2) Pipeline

4

## CONCLUSION

- 1) 모델 해석
- 2) 기대 효과
- 3) 활용 방안

# Context

1

## INTRO

- 1) 문제 이해
- 2) 문제 제기

2

## ANALYSIS

- 1) 도메인 분석
- 2) 데이터 분석
- 3) 시계열 한계
- 4) 데이터 가공

3

## MODEL

- 1) 사용 모델
- 2) Pipeline

4

## CONCLUSION

- 1) 모델 해석
- 2) 기대 효과
- 3) 활용 방안

## INTRO

- 문제 이해
- 문제 제기

## ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

## MODEL

- 사용 모델
- Pipeline

## CONCLUSION

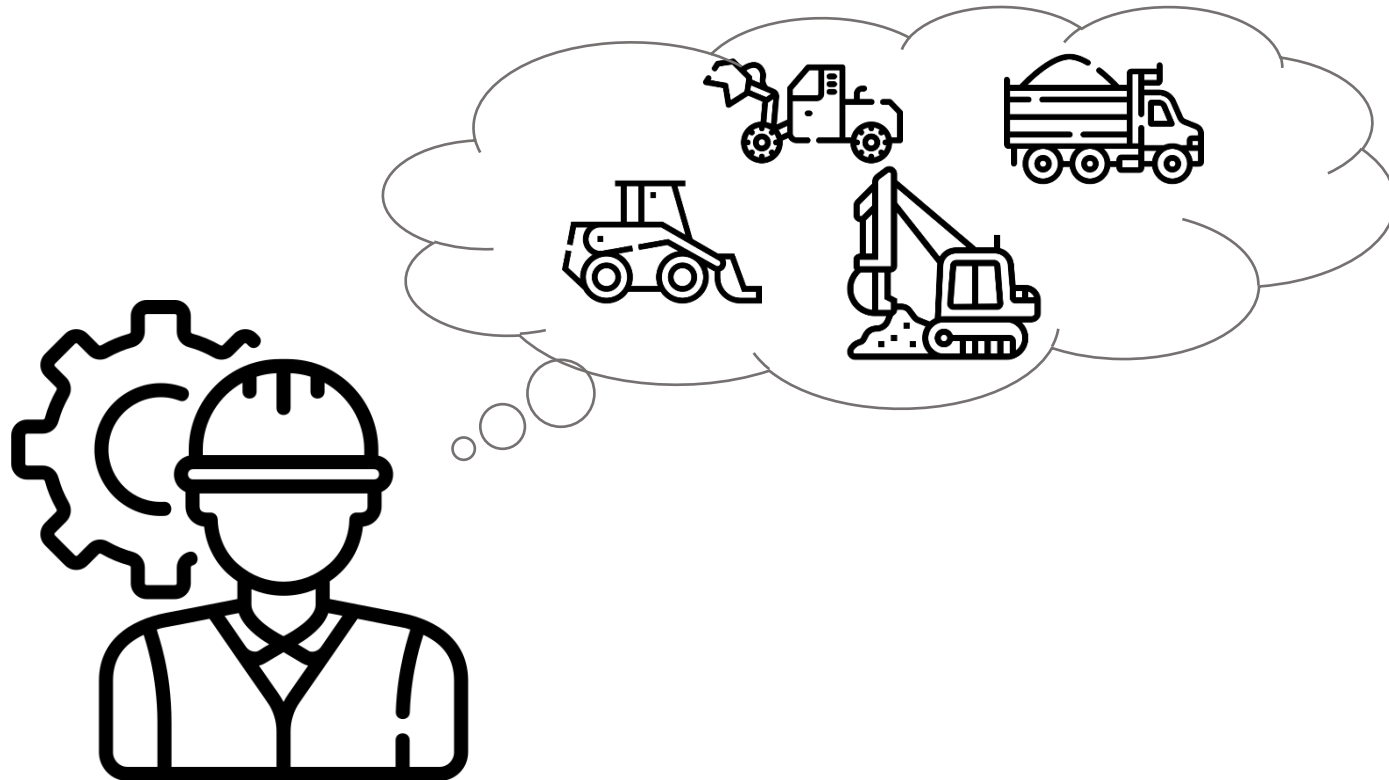
- 모델 해석
- 기대 효과
- 활용 방안

# Fleet Management 솔루션 필요성 대두

- Fleet management란 자동차, 지게차, 트레일러, 특수 차량과 같은 상업적 용도의 차량 관리를 의미함
- 효율적인 자원 관리는 비용을 절감하고 생산성을 향상시킴

## 최적화된 장비 조합

- 최적화된 장비 조합을 위해서는 **장비의 작업량**, 정비 기록 등을 통한 장비 정보가 필요함



## 문제 제기 – 작업자 영향력

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

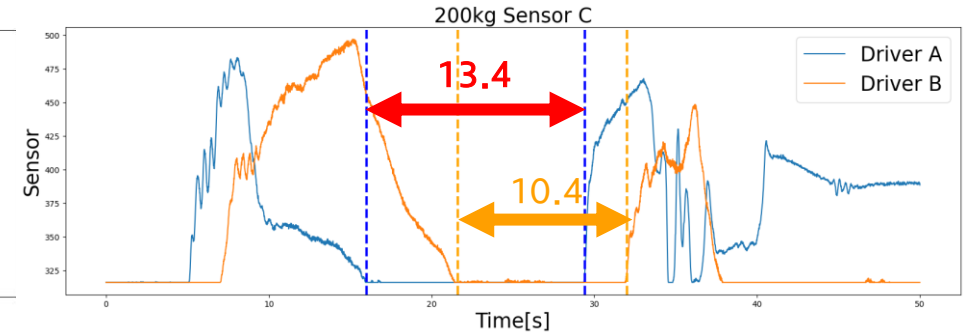
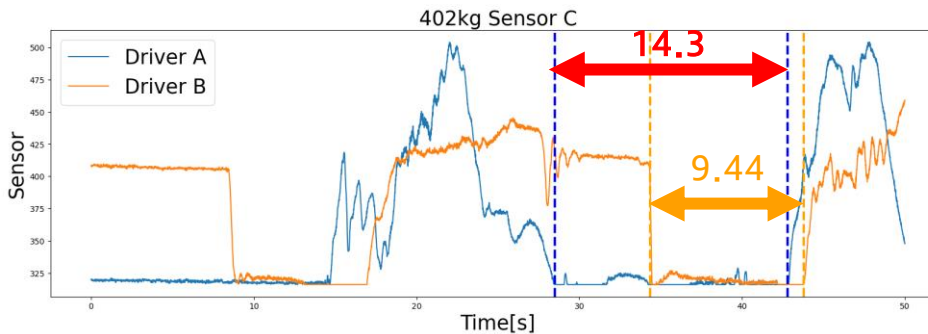
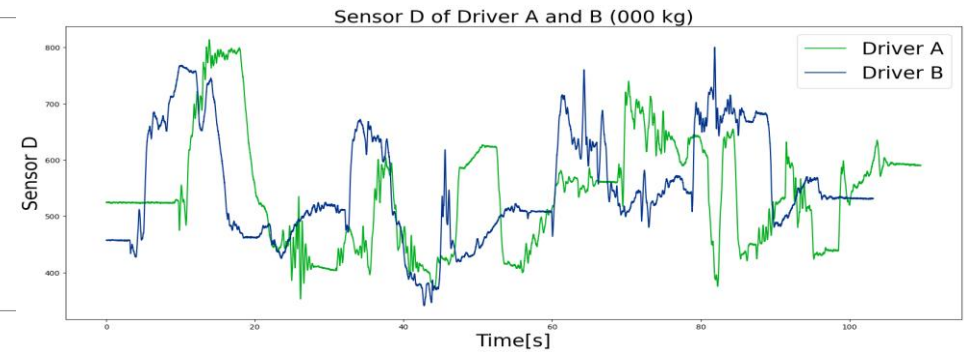
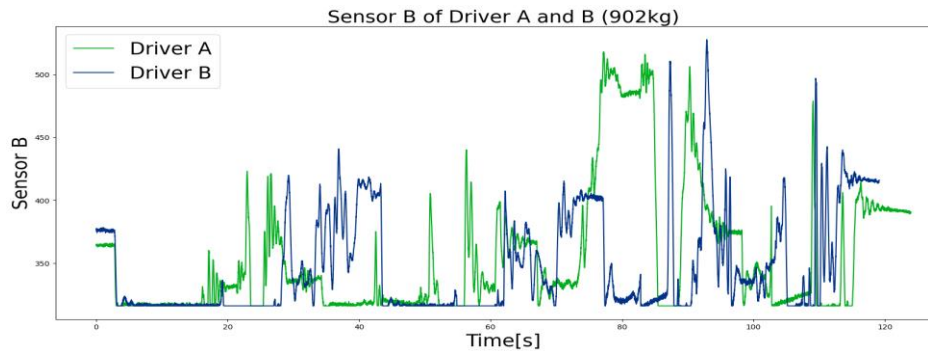
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- 작업자의 영향력이 강한걸 알 수 있음
  - 같은 중량과 Sensor여도 작업자의 따른 다른 형태의 행동을 보임
    - 모델의 Noise로 작용
    - 작업자 개개인에 영향을 받지 않는 모델 개발 필요



## 문제 제기 - Overpred

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

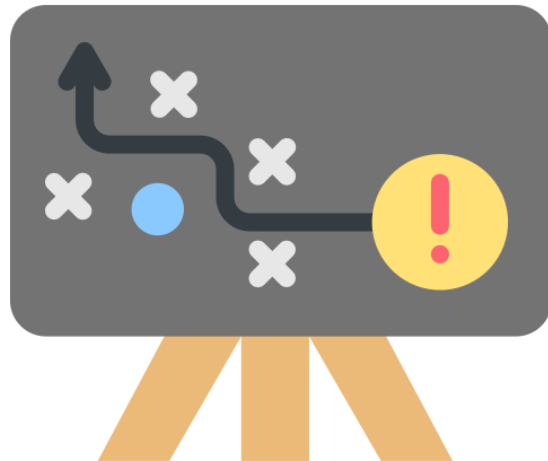
- 모델 해석
- 기대 효과
- 활용 방안

- 예측된 작업 중량이 실제 작업 중량보다 크면 발생하는 문제

1. 향후 남은 작업량이 더 많기 때문에 **건설 계획 수립 및 실행에 차질이 생김**  
이는 효율적인 현장 운용을 방해하고 산업 현장에서 굉장히 큰 위험 요소임

2. 장비 운용의 **비효율**이 발생함

앞선 예측 작업 중량이 실제 작업 중량보다 많다면 다음 장비는 실제로 필요한 것보다 더 많이 배치되게 되므로 장비를 낭비하게 됨



## 문제 제기 - Overpred

- 예측된 작업 중량이 실제 작업 중량보다 크면 발생하는 문제
  1. 향후 남은 작업량이 더 많기 때문에 건설 계획 수립 및 실행에 차질이 생김  
이는 효율적인 현장 운용을 방해하고 산업 현장에서 굉장히 큰 위험 요소임
  2. 장비 운용의 비효율이 발생함  
실제로 필요한 것보다 더 많이 배치되게 되므로 장비를 낭비하게 됨

**작업자의 운전 성향과 숙련도에 영향을 받지 않고**

**정확한 작업 중량을 예측하는 모델 개발**



- 문제 이해도
- 배경
- 문제제기
- 데이터 활용
- 도메인 분석
- 데이터 분석
- 시계열 분석
- 시계열 한계
- 모델 활용
- 사용 모델
- 평가
- 모델 해석
- 마무리
- 활용 방안
- 기대 효과

# Context

1

## INTRO

- 1) 문제 이해
- 2) 문제 제기

2

## ANALYSIS

- 1) 도메인 분석
- 2) 데이터 분석
- 3) 시계열 한계
- 4) 데이터 가공

3

## MODEL

- 1) 사용 모델
- 2) Pipeline

4

## CONCLUSION

- 1) 모델 해석
- 2) 기대 효과
- 3) 활용 방안



## 도메인 분석

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

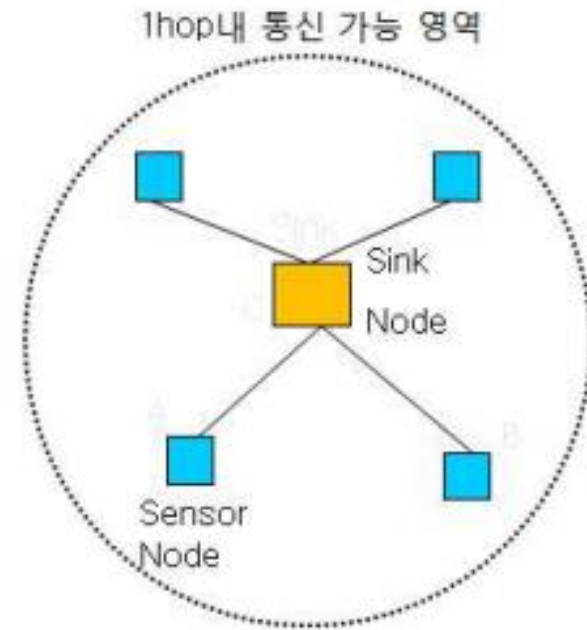
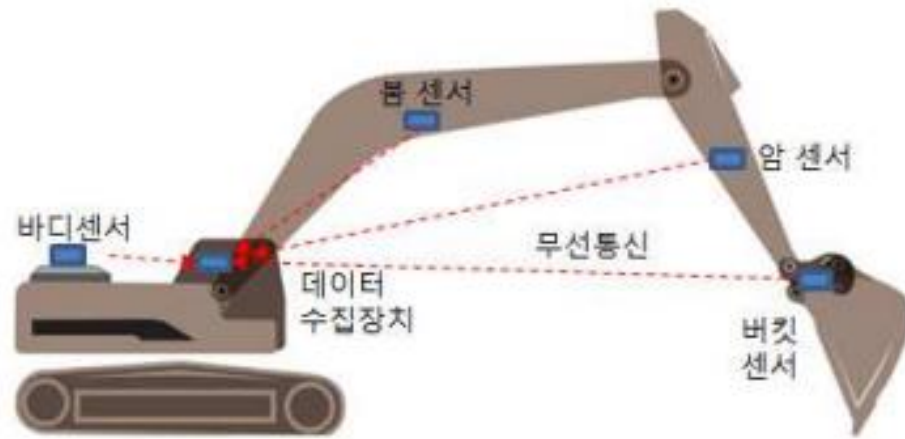
- 모델 해석
- 기대 효과
- 활용 방안

- 데이터의 특성과 관련된 건설 기계와 주최 회사 도메인 분석
- 다음과 같은 순서로 분석 진행



## 도메인 분석

- 데이터의 특성과 주최 회사 관련 도메인 분석
- 해당 데이터가 굴삭기 데이터이며, 4개의 Sensor라고 판단



[구성장치간 네트워크 Topology]

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

## 도메인 분석

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

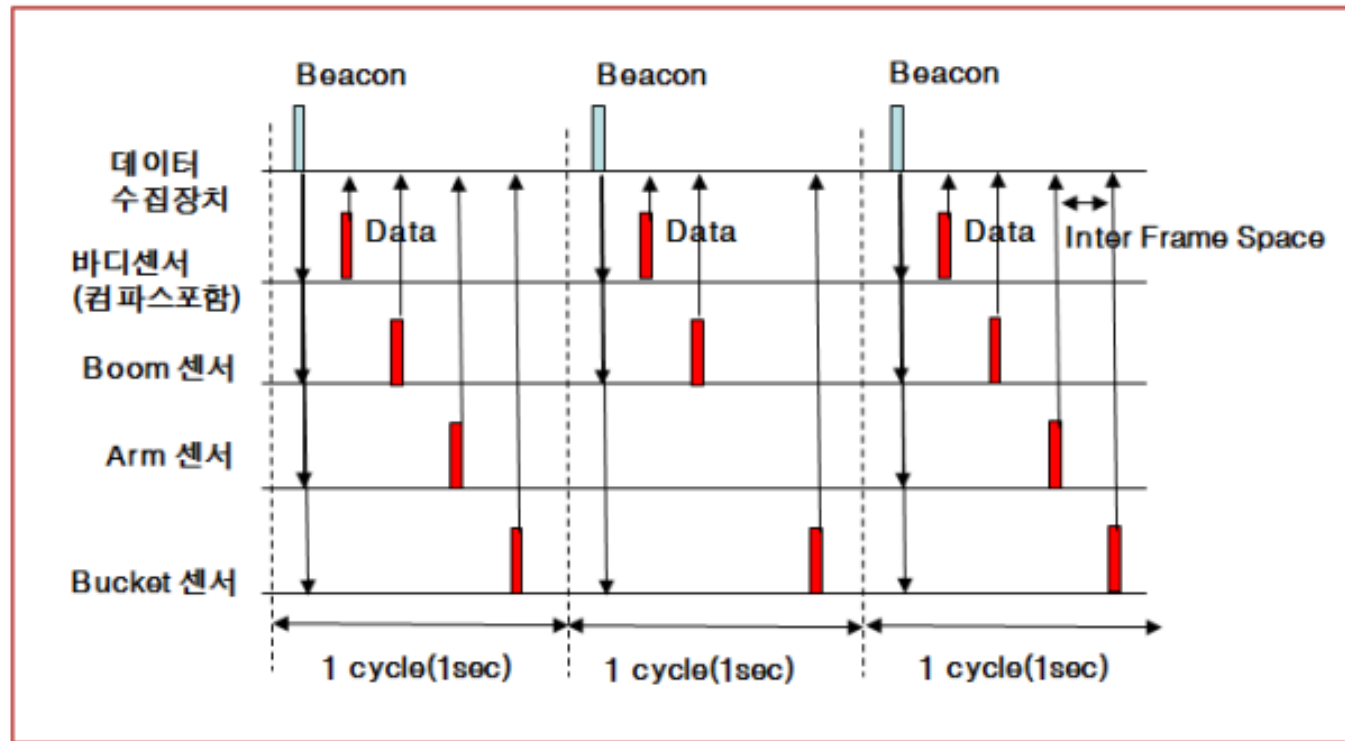
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- Sensor는 beacon을 통해 통신하는 것으로 예상
- 통신 순서는 Body – Boom – Arm – Bucket Sensor 순서로 통신



[채널 접근 및 데이터 송수신 절차]

## EDA Sensor Matching

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

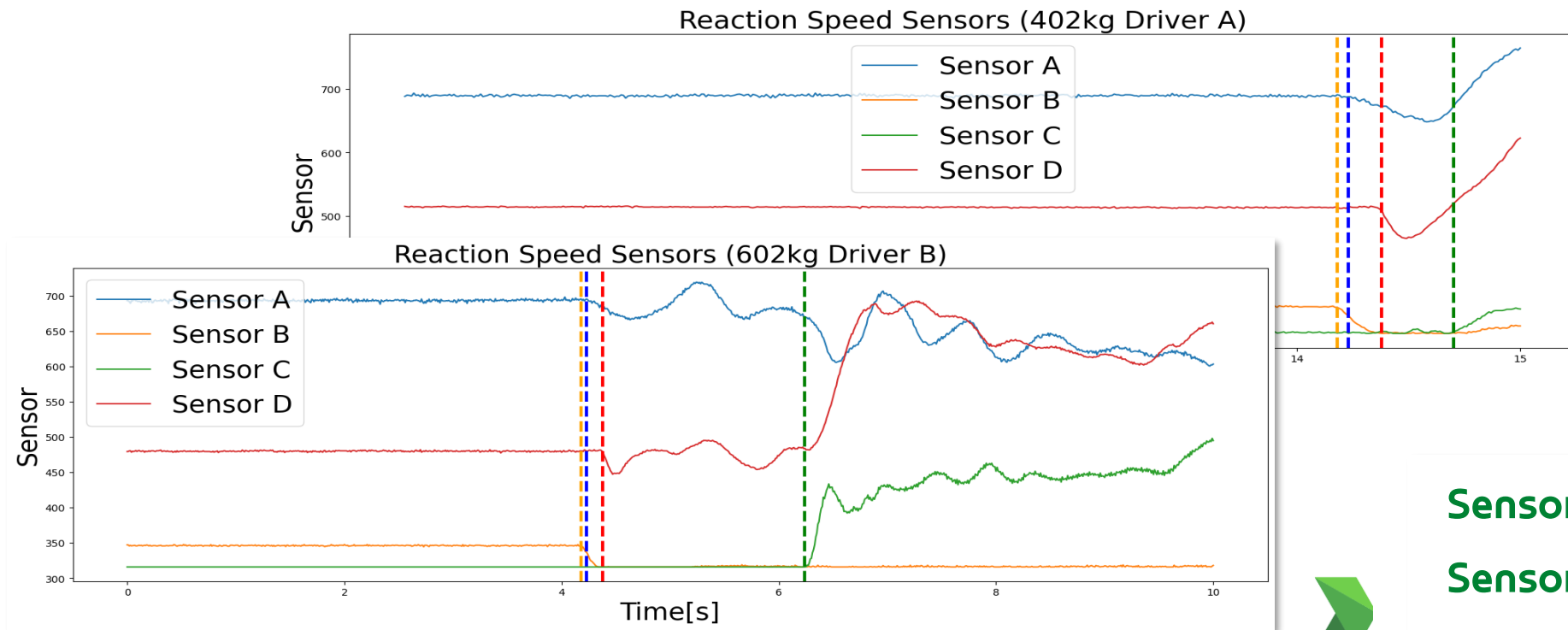
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- EDA를 통해 Sensor의 순서 차이를 확인
- 해당 검증을 통해 Sensor A, B, C, D 매칭



Sensor A → Boom

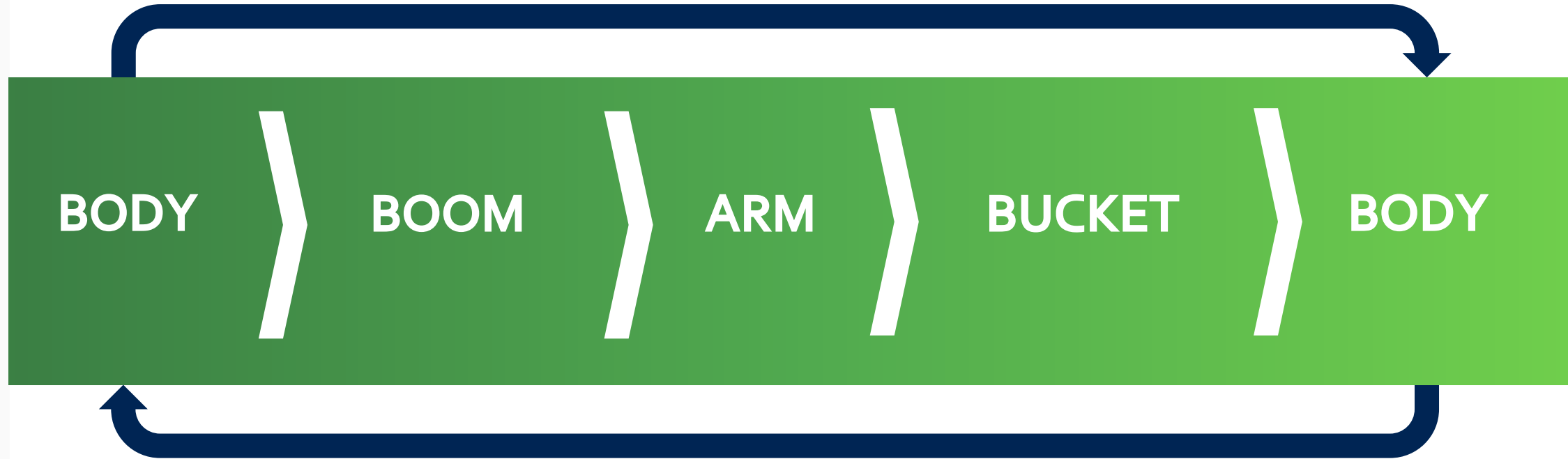
Sensor B → Body

Sensor C → Bucket

Sensor D → Arm

## 도메인 분석

- 포크레인은 보편적으로 땅을 파거나 물건을 집어서 위치를 이동시킴
- 위 행위를 위해서는 일련의 절차가 보편적으로 수행함
- 따라서 해당 순서로 작동하는 것이 보편적임



### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

## EDA Sensor Matching

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

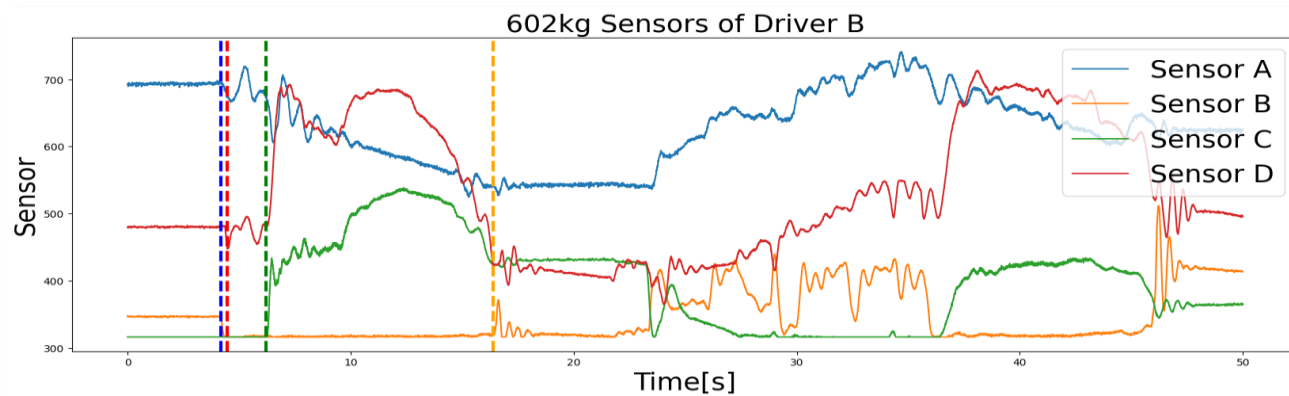
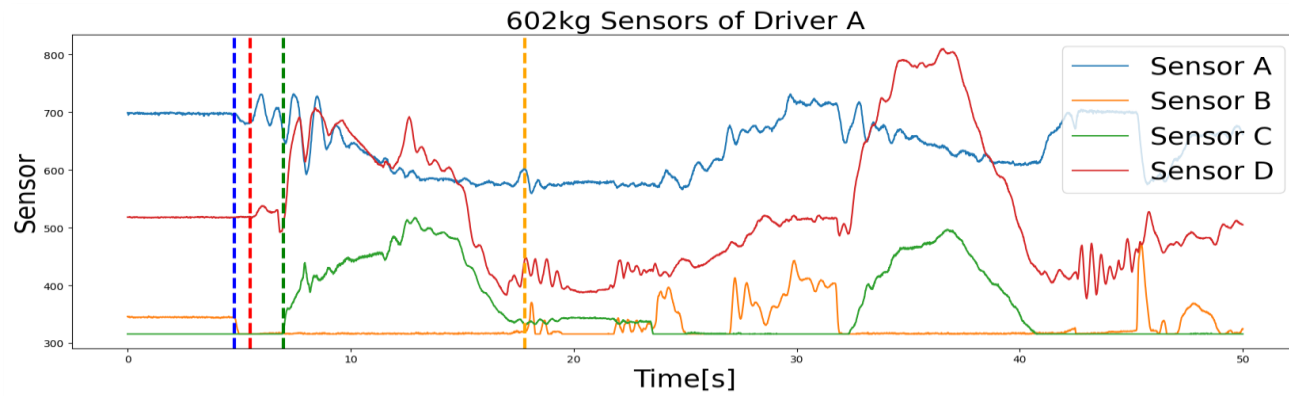
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

1. 원하는 방향에 도착함 → Boom이 움직임
2. Arm → Bucket이 움직임
3. Body가 움직임
4. 다시 Boom이 움직임 → 작업 순서에 해당하는 Sensor를 매칭함



**Sensor A → Boom**

**Sensor B → Body**

**Sensor C → Bucket**

**Sensor D → Arm**

## EDA 검증

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

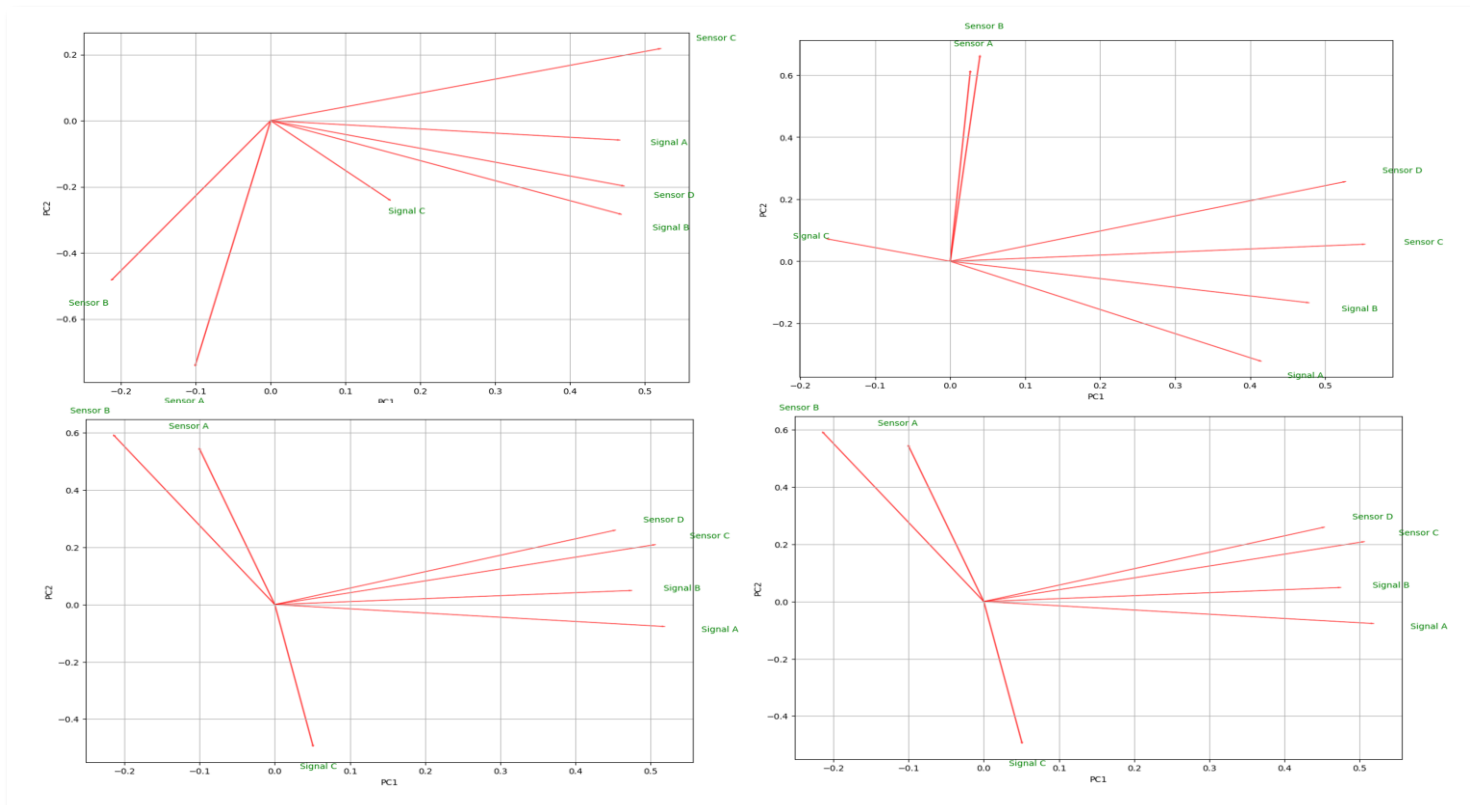
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- 앞서 매칭한 Sensor 데이터를 확인하기 위해 EDA 진행함
- Sensor A/B과 같은 경향 Sensor C/D가 같은 경향을 보임



## Insight

- 1) 압축된 차원을 X,Y축으로 설정
- 2) 두 개의 차원에 각 특징이 어떤 방향으로 영향을 주었는지 시각화
- 3) 같은 방향의 특성은 같은 영향을 미침

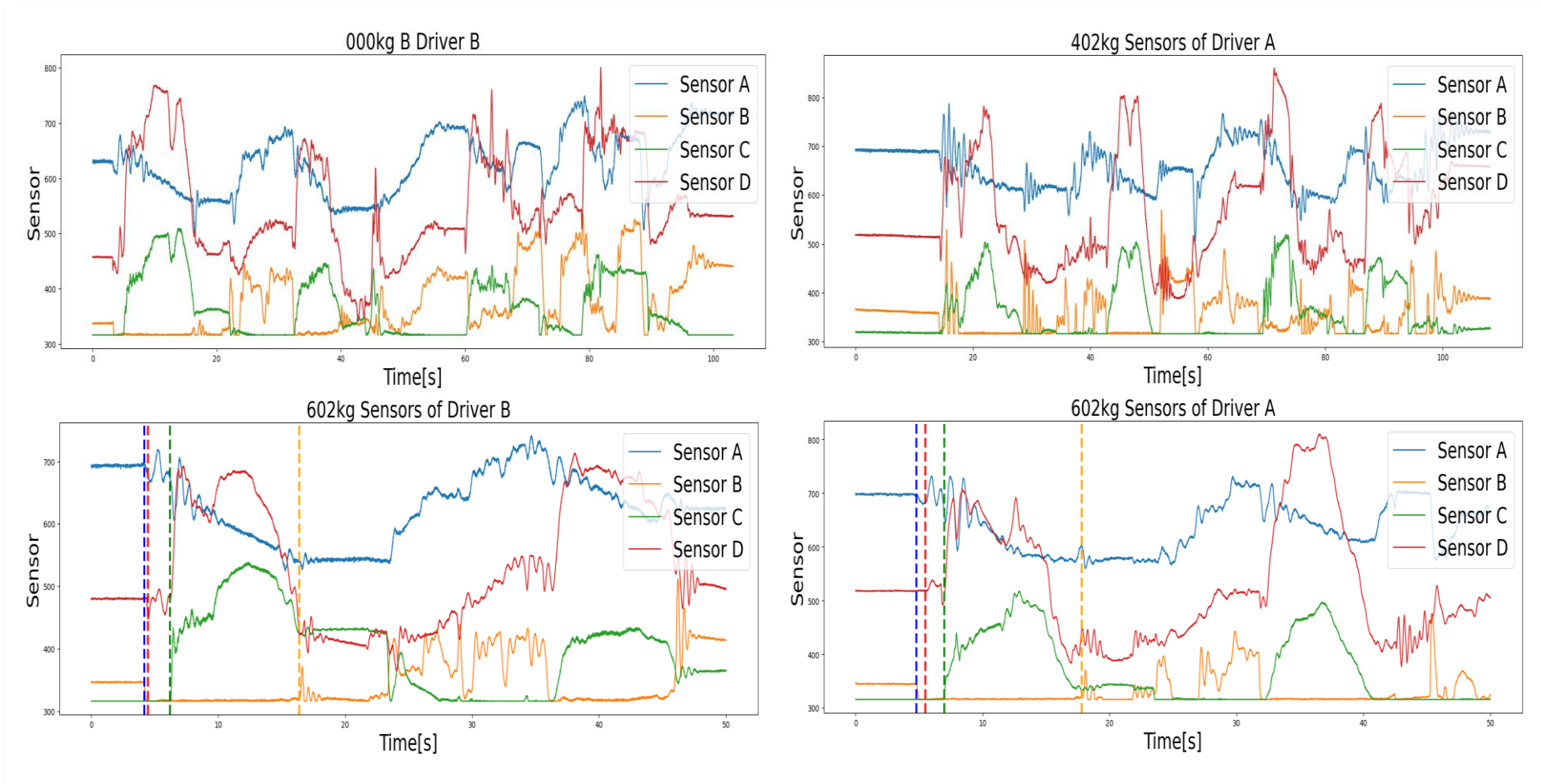
## Example

- 1) Sensor A/B가 같은 방향으로 시각화 됨
- 2) Sensor A/B가 같은 영향을 끼침을 알 수 있음

PCA: PCA란 다차원 데이터의 차원을 줄여 고유 주성분으로 데이터를 변환하여 데이터의 주요 정보를 추출하는 다변량 분석 및 차원 축소 기법

## 제공 데이터

- 비식별화 된 Signal 및 Sensor 시계열 데이터





## 시계열 모델의 한계

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

#### 짧은 Time Stamp 길이

- 시계열 모델은 시간에 걸쳐 나타나는 복잡한 종속성과 패턴을 학습하도록 설계
- 긴 시퀀스에서 학습한 패턴을 짧은 테스트 데이터에 효과적으로 일반화하는 데 어려움

#### 시작 시점의 변동성

- 시계열 모델은 시간적 컨텍스트를 기반으로 학습
- Test의 데이터셋의 시점은 언제일지 모르니 성능 저하

#### 부족한 훈련 데이터셋

- 시계열 모델은 훈련에서 학습한 패턴을 테스트에 일반화하기 위해 충분한 양의 학습 데이터셋이 필요

~~Time-Series~~

## 시계열 모델의 한계

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

#### 짧은 Time Stamp 길이

- 시간에 걸쳐 나타나는 복잡한 종속성과 패턴을 학습하도록 설계
- 긴 시퀀스에서 학습한 패턴을 짧은 테스트 데이터에 효과적으로 일반화하는 데 어려움

#### 시작 시점의 변동성

- 시계열 모델은 시간적 컨텍스트를 기반으로 학습
- Test의 데이터 시작 시점이 언제일지 모르니 성능 저하

#### 부족한 훈련 데이터셋

- 시계열 모델은 훈련에서 학습한 패턴을 테스트에 일반화하기 위해 충분한 양의 훈련 데이터셋이 필요

# Machine Learning

~~Time-Series~~

## Machine Learning Model

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

#### 짧은 Time Stamp 길이

- 머신러닝 모델은 시간의 변화보다는 각 타임스탬프에서의 특성 값들의 관계에 중점
- 각 샘플 내의 특성 패턴을 학습하여 짧은 시간 내의 중요한 정보를 효과적으로 예측하는 데 유리

#### 시작 시점의 변동성

- 각 샘플을 시간적 컨텍스트와 독립적으로 처리
- 어느 시점에서 샘플링 되었는지에 관계없이 일관된 예측을 제공

#### 부족한 훈련 데이터셋

- 머신러닝 모델은 주어진 특성들 간의 복잡한 관계를 학습하는데 중점
- 훈련 데이터셋이 작을 때 유리

**Machine Learning 활용**

## 효과적인 학습을 위한 데이터 가공

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

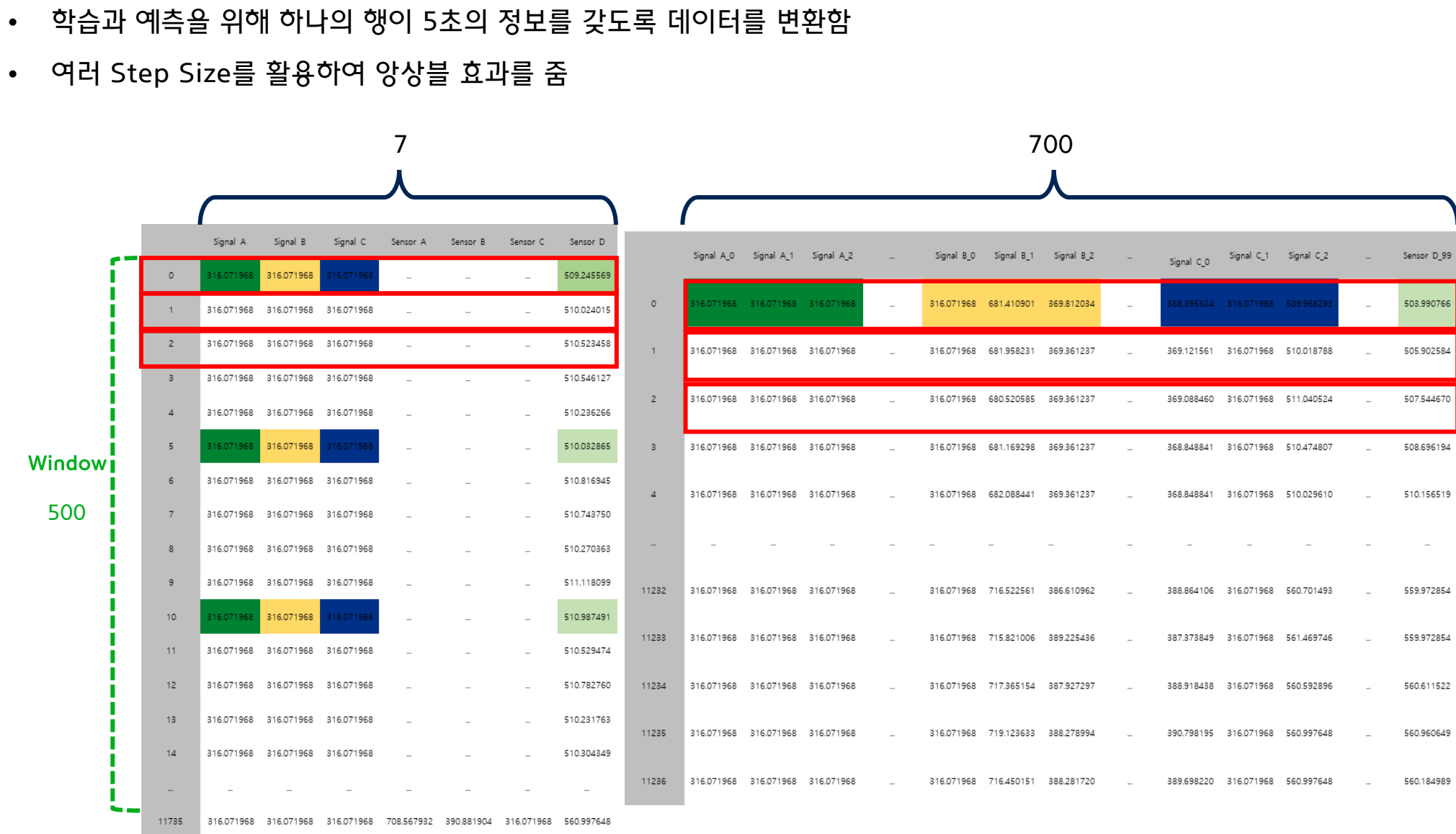
- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안



## 작업자에 영향 받지 않는 데이터 가공

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

1. 운전자의 숙련도에 따라 동시에 조작하는 사람, 시간이 걸리는 사람이 존재
2. 숙련도의 차이가 있더라도 5초 안에 하고자 하는 행위/조작을 할거라 가정
3. 해당 부분을 반영하기 위해 기존 데이터를 윈도우 500으로 설정하여 한 행에 반영되도록 transform 진행
4. 생성된 데이터의 한 행에는 운전자의 5초 정보가 포함되어 있음
5. 이를 통해 중량을 예측하는데 운전자의 5초의 정보가 들어감으로서,  
운전자의 숙련도에 영향 받지 않고 정확한 중량 측정 가능

7

Window  
500

700

	Signal A	Signal B	Signal C	Sensor A	Sensor B	Sensor C	Sensor D
0	316.071968	316.071968	316.071968	—	—	—	509.245569
1	316.071968	316.071968	316.071968	—	—	—	510.024015
2	316.071968	316.071968	316.071968	—	—	—	510.523458
3	316.071968	316.071968	316.071968	—	—	—	510.546127
4	316.071968	316.071968	316.071968	—	—	—	510.236266
5	316.071968	316.071968	316.071968	—	—	—	510.032865
6	316.071968	316.071968	316.071968	—	—	—	510.816945
7	316.071968	316.071968	316.071968	—	—	—	510.743750
8	316.071968	316.071968	316.071968	—	—	—	510.270363
9	316.071968	316.071968	316.071968	—	—	—	511.118099
10	316.071968	316.071968	316.071968	—	—	—	510.987491
11	316.071968	316.071968	316.071968	—	—	—	510.529474
12	316.071968	316.071968	316.071968	—	—	—	510.782760
13	316.071968	316.071968	316.071968	—	—	—	510.231763
14	316.071968	316.071968	316.071968	—	—	—	510.304349
...	...	...	...	...	...	...	...
11735	316.071968	316.071968	316.071968	708.567932	390.881904	316.071968	560.997648

	Signal A_0	Signal A_1	Signal A_2	...	Signal B_0	Signal B_1	Signal B_2	...	Signal C_0	Signal C_1	Signal C_2	...	Sensor D_99
0	316.071968	316.071968	316.071968	—	316.071968	681.410901	369.812034	—	368.998324	316.071968	509.963299	—	509.990766
1	316.071968	316.071968	316.071968	—	316.071968	681.958231	369.361237	—	369.121561	316.071968	510.018788	—	505.902584
2	316.071968	316.071968	316.071968	—	316.071968	680.520585	369.361237	—	369.088460	316.071968	511.040524	—	507.544670
3	316.071968	316.071968	316.071968	—	316.071968	681.169298	369.361237	—	368.848841	316.071968	510.474807	—	508.696194
4	316.071968	316.071968	316.071968	—	316.071968	682.088441	369.361237	—	368.848841	316.071968	510.029610	—	510.156519
...	...	...	...	...	...	...	...	...	...	...	...	...	...
11232	316.071968	316.071968	316.071968	—	316.071968	716.522561	386.610962	—	388.864106	316.071968	560.701493	—	559.972854
11233	316.071968	316.071968	316.071968	—	316.071968	715.821006	389.225436	—	387.373849	316.071968	561.469746	—	559.972854
11234	316.071968	316.071968	316.071968	—	316.071968	717.365154	387.927297	—	388.918438	316.071968	560.592896	—	560.611522
11235	316.071968	316.071968	316.071968	—	316.071968	719.123633	388.278994	—	390.798195	316.071968	560.997648	—	560.960649
11236	316.071968	316.071968	316.071968	—	316.071968	716.450151	388.281720	—	389.698220	316.071968	560.997648	—	560.184989

## Feature engineering

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- 운전자의 스타일과 숙련도가 모델의 예측에 미치는 영향을 줄이고, 객관적인 작업 중량 예측을 위해 파생변수 생성 필요함

#### 부하 토크

- Boom과 Arm의 움직임으로 발생하는 힘의 회전적 효과를 나타냄
- 굴삭기에서는 통상적으로 Boom, Arm이 물체를 들어올리거나 이동시키는 데 필요한 힘과 관련됨
- 각도 변화율로 계산할 수 있으며, 굴삭기의 작업 중량을 추정하는 데 도움을 줄 수 있음

#### 동적 반응

- 센서 값의 변화율의 변화율, 즉 가속도를 나타냄
- 센서의 변화가 얼마나 빠르게 가속 또는 감속되는지를 나타내며, 동적 부하의 변화를 측정하는 데 유용함
- 동적 반응은 작업 중량의 변화에 따라 달라질 수 있으며, 중요한 정보를 제공할 수 있음

# Context

1

## INTRO

- 1) 문제 이해
- 2) 문제 제기

2

## ANALYSIS

- 1) 도메인 분석
- 2) 데이터 분석
- 3) 시계열 한계
- 4) 데이터 가공

3

## MODEL

- 1) 사용 모델
- 2) Pipeline

4

## CONCLUSION

- 1) 모델 해석
- 2) 기대 효과
- 3) 활용 방안

## Modeling – AutoGluon-Tabular

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

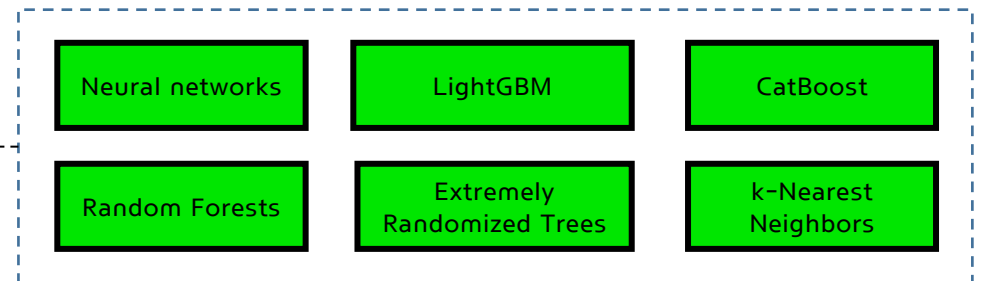
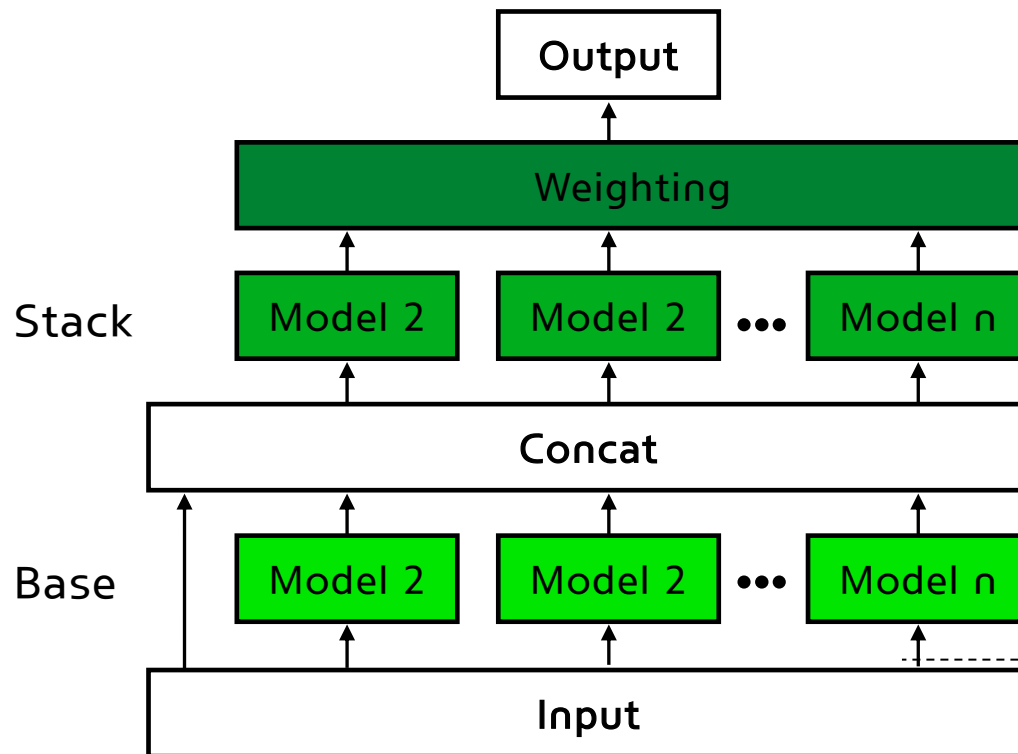
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

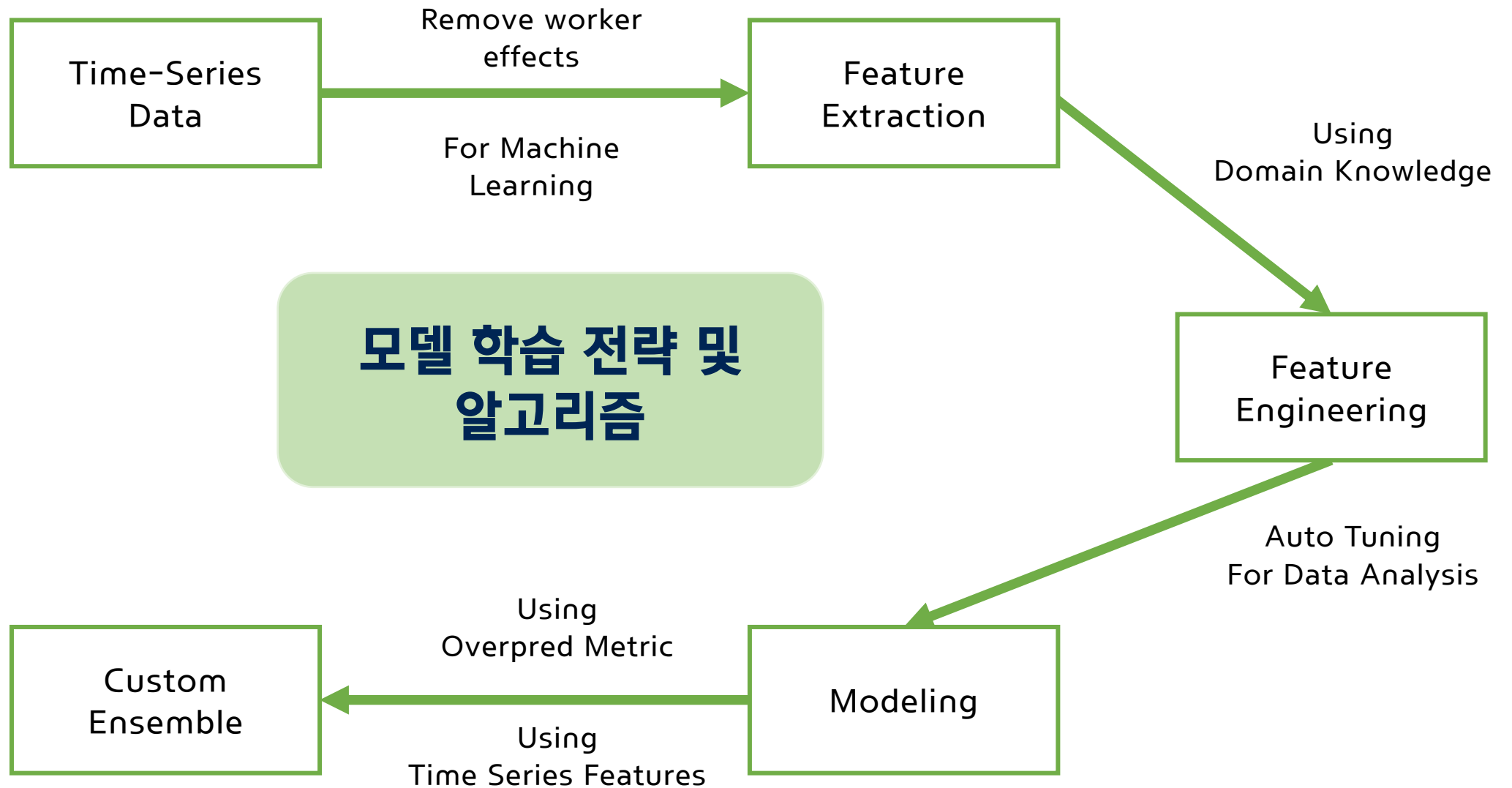
- 모델 해석
- 기대 효과
- 활용 방안

- AutoML 프레임워크
- 보편적인 AutoML이 주로 사용하는 CASH이 아닌, **ensembling**과 **stacking** 활용함
- 각 모델 학습 시 Repeated k-fold ensemble bagging 사용함
  - 이를 통해 과적합을 방지하고자 함





## Overall Pipeline



### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

# Context

1

## INTRO

- 1) 문제 이해
- 2) 문제 제기

2

## ANALYSIS

- 1) 도메인 분석
- 2) 데이터 분석
- 3) 시계열 한계
- 4) 데이터 가공

3

## MODEL

- 1) 사용 모델
- 2) Pipeline

4

## CONCLUSION

- 1) 모델 해석
- 2) 기대 효과
- 3) 활용 방안

## 모델 해석

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

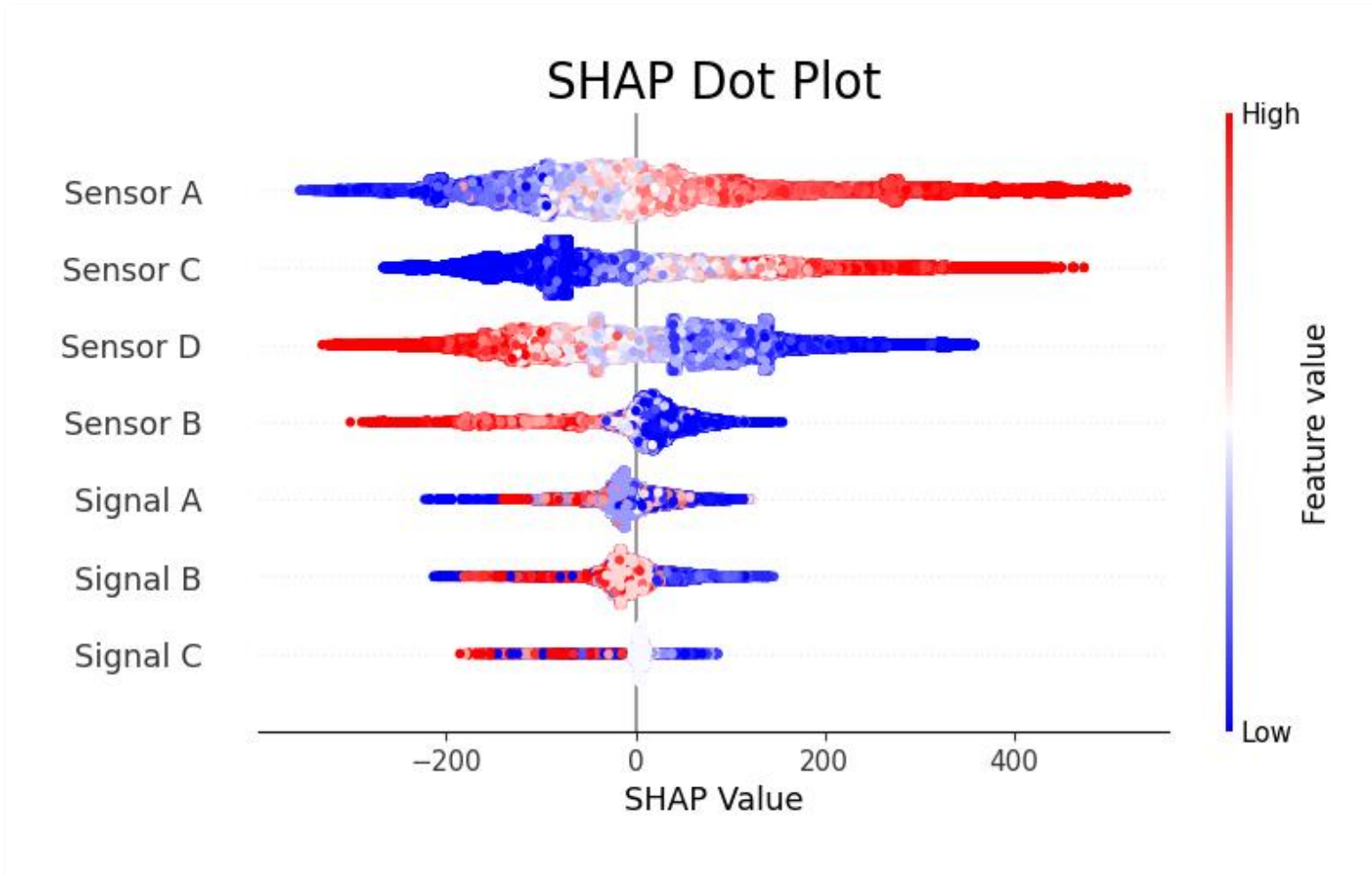
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- Sensor A와 C가 같은 경향성을 B와 D는 반대의 경향성을 가짐
- Sensor A의 값이 높게 나올 수록 중량이 높은 걸로 확인 됨



## 모델 해석

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

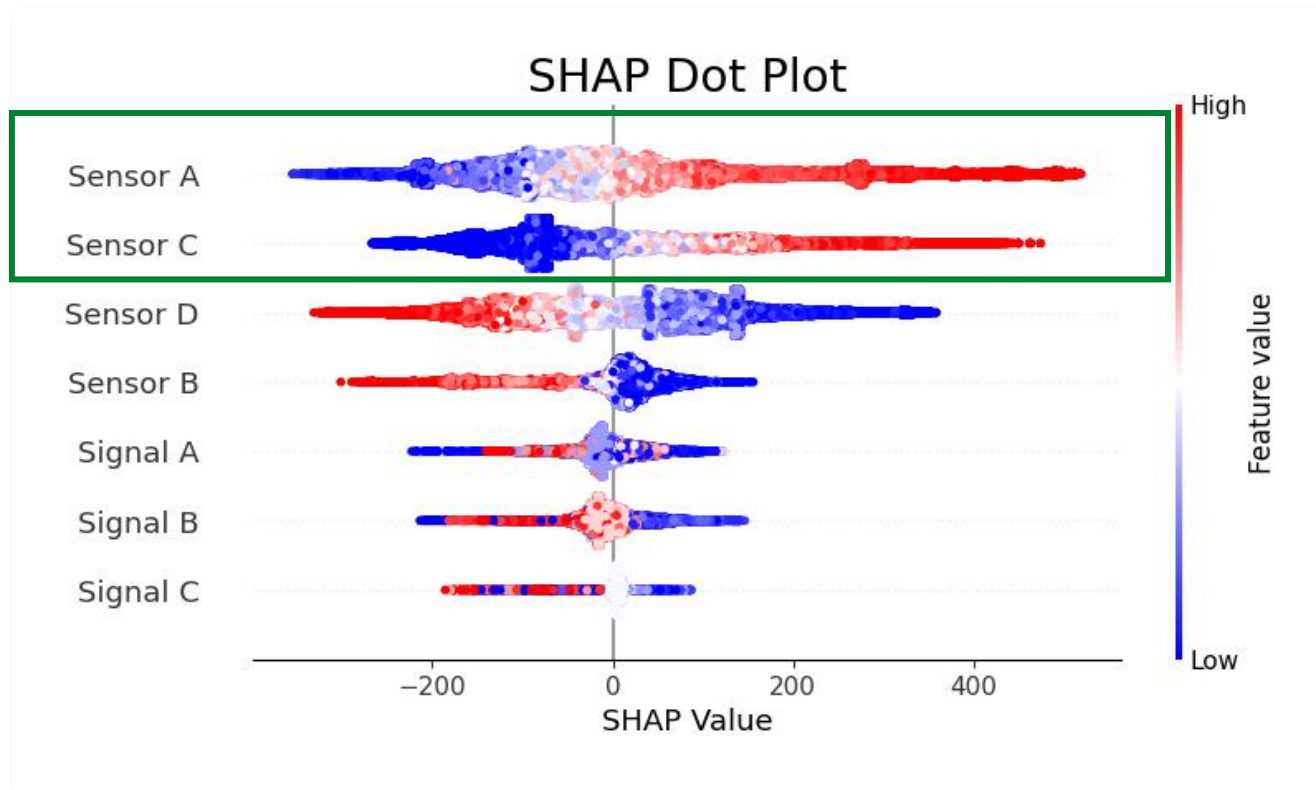
### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

- 모델 해석
- 기대 효과
- 활용 방안

- Sensor A와 C는 Boom과 Bucket, 땅을 파거나 물건을 들어올릴 때 직접적인 영향을 받는 Sensor
- Sensor A와 C가 값이 높게 나올 수록 중량이 높게 측정됨



### Insight

- 1) 해당 모델을 통한 추가적인 분석도 가능
- 2) 센서 데이터 이상치 탐지와 같은 활용 방안

### Example

- 1) 굴삭기가 가변운 중량을 들었음에도 불구하고 Sensor A에서 높은 값이 나온다면 이는 기계 이상 신호를 보임
- 2) 분석 모델의 넣지 않더라도, 현장에서 작업자가 Sensor의 값으로도 기기 이상 및 정확한 중량이 측정되는지 확인 가능

## 기대 효과 & 활용 방안

### INTRO

- 문제 이해
- 문제 제기

### ANALYSIS

- 도메인 분석
- 데이터 분석
- 시계열 한계
- 데이터 가공

### MODEL

- 사용 모델
- Pipeline

### CONCLUSION

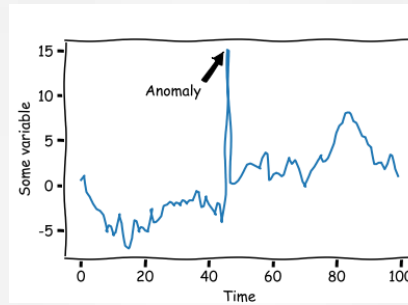
- 모델 해석
- 기대 효과
- 활용 방안



최적화된 장비 조화

작업 효율 극대화

센서 데이터를 활용한 정확한 중량 확인  
이로 인한 작업 효율 극대화



이상 탐지

안전성 확보

중량 예측을 통한 센서 데이터 분석  
이로 인한 안정성 확보



자동 굴착기 가능

AI를 통한 자동화

현대 건설기계 자동 굴착기에 사용 가능  
AI를 통한 자동 굴착기 생성

**THANK YOU**