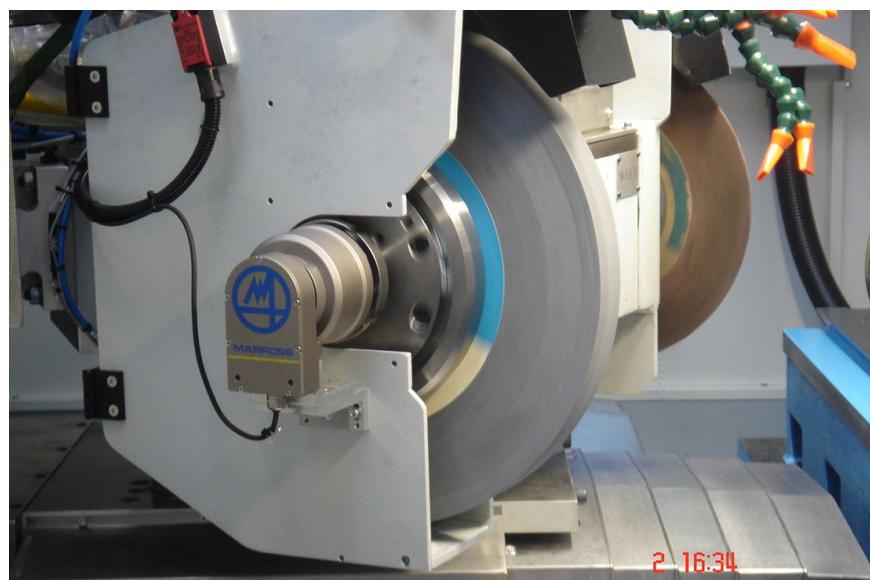


<AI을 이용한 연삭 훈수명 관리 및 교체시기 예측 솔루션>

프로젝트 개발서



[업타운 Team Member]

최창규(팀장)

강보원

박현명

최정인

| Index

1. 프로젝트 개요

- 1.1. 프로젝트 배경 및 목적
- 1.2. 프로젝트 개발환경
- 1.3. 프로젝트 작업 프로세스

3. 홈페이지 설명

- 3.1. 데이터 대시보드 상세설명

5. 기대효과

- 5.1. 기대효과

7. 개선점 및 발전 사항

- 7.1. 개선점
- 7.2. 발전사항

2. 데이터 설명

- 2.1. 데이터설명
- 2.2. 데이터 구축 구성도
- 2.3. 데이터 수집 대상 장비
- 2.3. 데이터 축적 현황

4. 알고리즘

- 4.1. 사용 알고리즘
- 4.2. 적용 결과 - PoC 결과 후

6. 프로젝트 결과 보고

- 6.1. 데이터 전처리
- 6.2. 딥러닝
- 6.3. 머신러닝
- 6.4. DB설계
- 6.5. 웹 플랫폼

1. 프로젝트 개요

1.1 프로젝트 배경 및 목적

1.1.1 프로젝트 주제 및 배경

[연삭기 센서 데이터 및 MES 분석을 통한 연삭 훈 수명관리 및 교체시기 예측 솔루션]

연삭기 센서 데이터 및 MES 데이터를 분석하여 실시간으로 연삭 훈의 수명을 관리하고 교체 시기를 예측하여 설비 운영 효율화 및 제품 고도화를 위해 AI 솔루션 도입이 필요함

1.1.2 프로젝트 목적

- 수요기업에는 스마트 공장 구축사업을 통해 기본적인 스마트 공장 구축을 위한 준비를 하였으나, 제조 설비가 동작할 때 수집되는 빅데이터를 활용하지 못하는 문제가 존재함.
- 지속적으로 품질 향상에 대한 요구사항이 증가하고 있으며 원활한 설비 가동을 통한 생산성 및 품질 향상을 위해 설비에서 발생하는 데이터를 AI 와 융합하여 설비의 건전성을 파악하고자 함

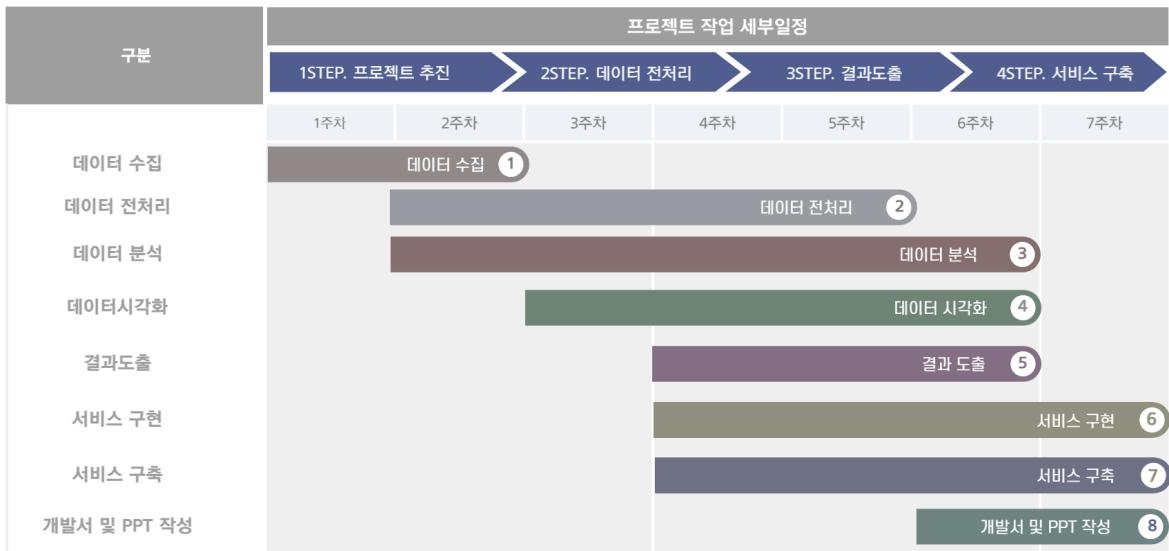
1.2 프로젝트 개발환경

1.2.1 개발환경

개발환경	<ul style="list-style-type: none">▶ OS (운영체제) 버전 : Windows 10 / Windows 11▶ 개발툴 : Colab pro / visual studio code▶ 데이터 베이스 : influxdb 1.8.10▶ 형상 관리 : Git, GitHub ▶ 데이터 분석 버전: Python 3.9.7▷ 전처리: Numpy 1.23.0, Pandas 1.4.3▷ 시각화 : Grafana 9.3.2▷ 머신러닝: Keras 2.9.0, Tensorflow 2.9.2, Scikit-Learn 1.0.2▷ 딥러닝: ARIMA, LSTM ▶ 웹 개발 프레임워크 : Django 4.1.5▷ 프론트엔드: HTML5, JS, CSS
------	--

1.3 프로젝트 작업 프로세스

1.3.1 프로젝트 작업 세부일정



1.3.2 프로젝트 팀원 역할

편의를 세상에 뿌리다

UPTOWN

PROFILE	PROFILE	PROFILE	PROFILE
NAME 강보원 (1998.06.01) E-mail iambowonkang@gmail.com	NAME 박현명 (1995.02.25) E-mail wisdom01636@naver.com	NAME 최정인 (1998.05.17) E-mail junginmay17@naver.com	NAME 최창규 (2000.02.10) E-mail kimky944@gmail.com

- 강보원: 데이터 분석 및 딥러닝
- 박현명: 데이터 분석 및 시각화
- 최정인: 데이터 분석 및 웹 프로그래밍
- 최창규: 데이터 분석 및 프로젝트 총괄

1.3.3. 업무 세부사항

ⓐ 데이터 분석 및 모델링

1. 데이터 값 변화 확인, 각 열의 상관성 확인, 결측치 확인
2. 데이터의 필요없는 열 삭제(값 변화 없음, 상관성 없음, 결측치 다수)
3. 7 : 1로 train, test set를 생성 후 ARIMA, 7 : 3으로 LSTM에 적용

ⓑ 웹 페이지 작업

업무	업무 내용
DB 작업	- DB에 전처리 된 데이터를 저장 후 grafana에서 사용할 수 있게 제작
페이지 작업	- 홈페이지 : 페이지 소개, project소개 - login / signup 페이지 : 로그인과 회원가입 - 추가 필요

ⓒ 개발 Tool 활용

개발 Tool	업무 내용
Visual Studio Code	<ul style="list-style-type: none">• Django를 이용하여 웹 페이지 생성• DB 연동• JQuery,Css 활용하여 페이지 디자인• 웹 페이지 실행 시 오류 Debugging• 데이터 전처리

InfluxDB	<ul style="list-style-type: none">• DB 생성 및 grafana 연동
----------	--

2. 데이터설명

2.1 데이터설명

2.1.1 Edge PC로 수집되는 데이터

- 전류 센서 1개
- 진동 센서 1개
- Edge PC 상태

2.1.2 FANUC 데이터

- 휠 잔량
- 드레싱 주기
- 동작 상태

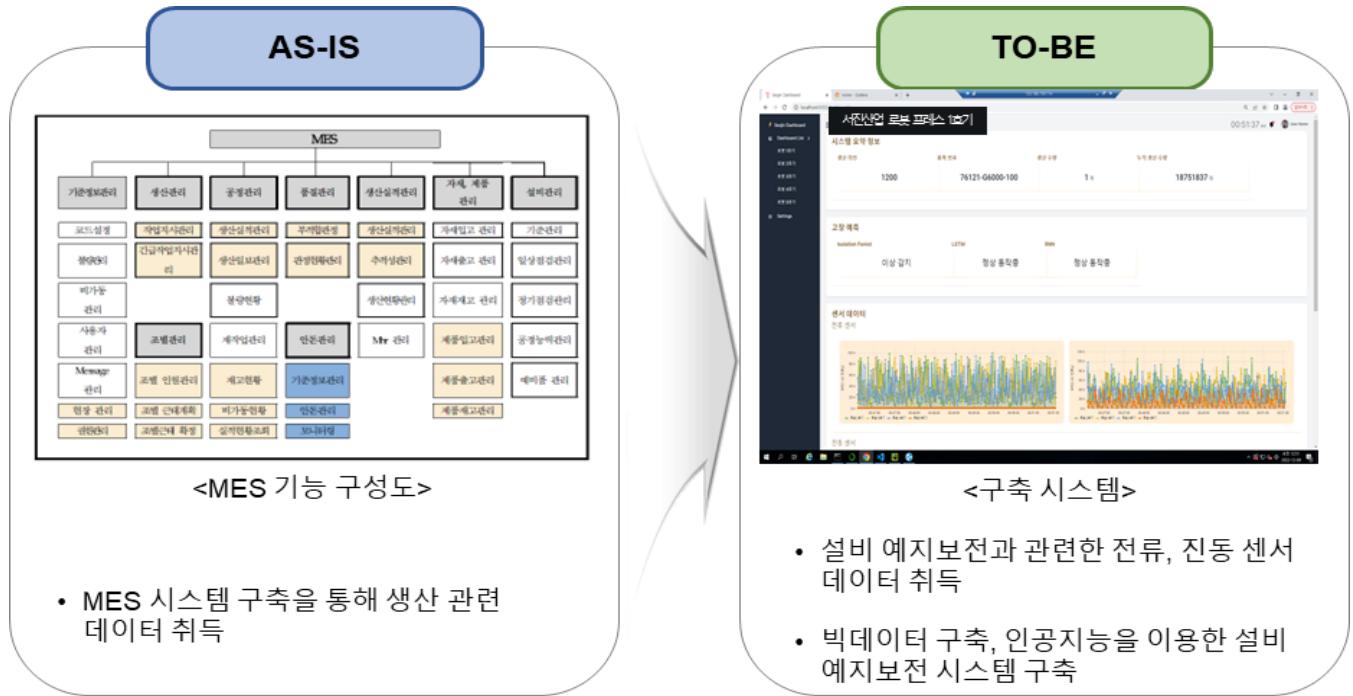
2.1.3 MES 데이터

- 생산 실적
- 재고 현황
- 불량 판정

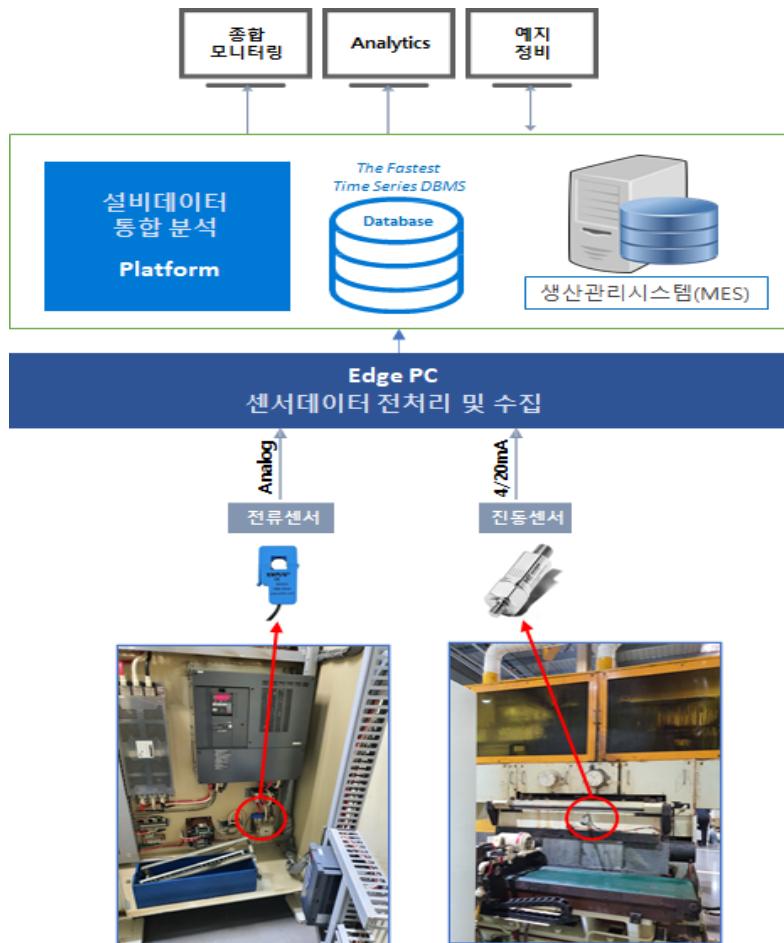
2.2 데이터 구축 구성도

- 제조 현장 정보를 실시간으로 확보함
- 현재는 과거에 비해 IoT 기술이 많이 발달하여 데이터를 수집하고 축적하기가 쉬움
- 설비에서 데이터를 직접 수집하거나 센서를 설치하여 생산성에 미치는 요소들을 수집하여 분석함
- 연삭기에 진동 센서, 그리고 분전함에 전류 센서를 설치하여 Edge PC에서 센서 데이터를 수집함
- 센서 데이터와 MES 데이터를 기반으로 함수식 계산, ARIMA, LSTM 모델을 이용하여 실시간으로 교체 시기 예측을 진행함
- MES, 진동, 전류, FANUC 데이터 수집함
- MES 데이터를 통해 생산 실적, 재고 현황, 불량 판정 등의 생산 이력과 관련한 데이터를 실시간으로 수집하고 연삭기에 설치한 진동과 전류 및 FANUC 데이터를 취합하여 시계열 DB를 구축함
- 설비에 Edge PC를 이용하여 센서 데이터를 1차로 수집하고 일정 주기마다 서버에 저장하도록 DB를 구축함

2.2.1 MES 기능 구성도 & 구축 시스템



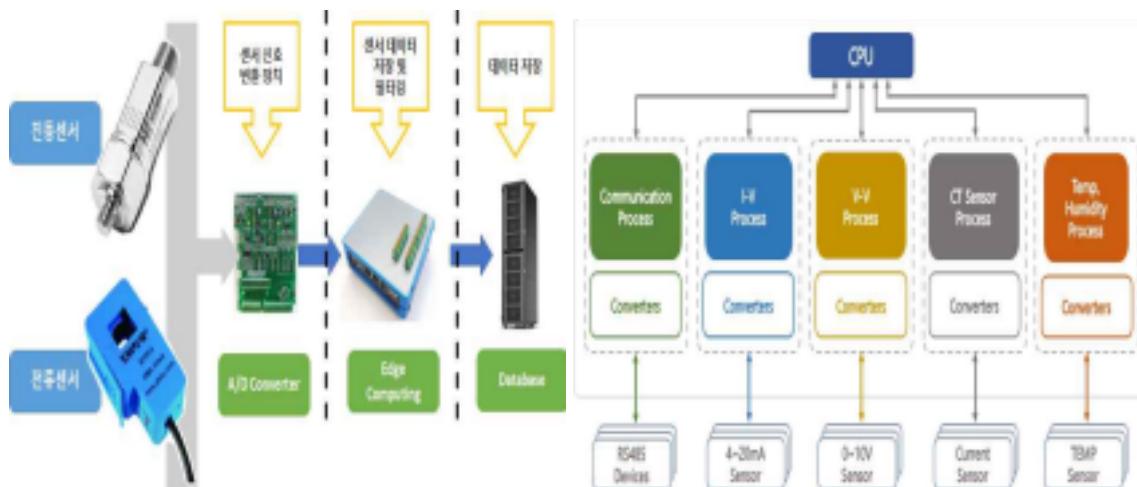
2.2.2 시스템 구성도



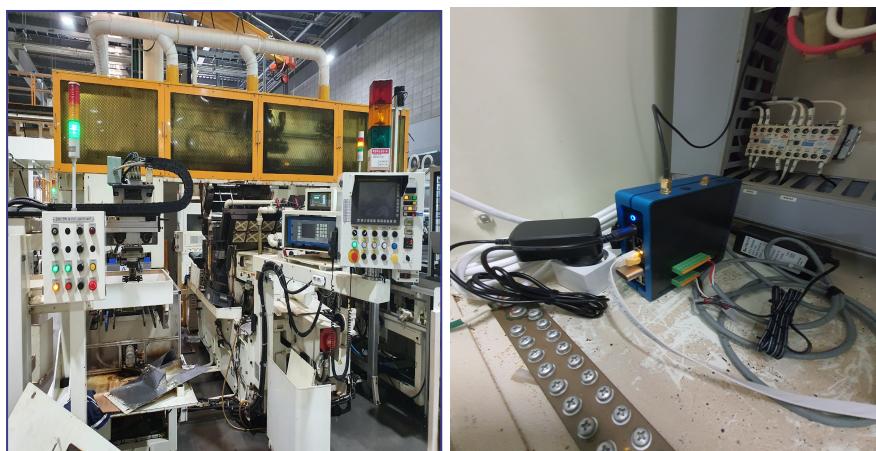
2.2.3 하드웨어 & 소프트웨어 구성도



2.2.4 센서 데이터 수집 & Edge PC 내부 인터페이스 구성도



2.3 데이터 수집 대상 장비



- 센서가 부착 될 연삭기 장비
- 센서 데이터가 수집 되는 Edge PC

2.4 데이터 축적 현황

2.4.1 Edge PC로 수집되는 데이터

name: Data	time	Irsm1	Irsm2	Irsm3	currvolt1	currvolt2	currvolt3	currvolt4	humidity	input1	input2	input3	input4	temp1	temp2	temperature
1668410976963996000	0.07	0.08	44.26	0.03	4.31	0	0.48	20.2	0	0	0	0	0	0	0	32.26
1668410975965025000	0.06	0.09	44.8	0.03	4.3	0	0.47	20.22	0	0	0	0	0	0	0	32.24
1668410974965111000	0.05	0.1	42.55	0.03	4.3	0	0.47	20.2	0	0	0	0	0	0	0	32.25
1668410973965210000	0.06	0.11	44.64	0.03	4.3	0	0.47	20.2	0	0	0	0	0	0	0	32.26
1668410972966314000	0.05	0.11	41.45	0.03	4.31	0	0.48	20.2	0	0	0	0	0	0	0	32.26
1668410971966450000	0.08	0.11	43.98	0.03	4.32	0	0.5	20.2	0	0	0	0	0	0	0	32.25
1668410970966665000	0.05	0.1	46.52	0.03	4.32	0	0.5	20.2	0	0	0	0	0	0	0	32.25
1668410969966734000	0.06	0.12	44.53	0.03	4.31	0	0.48	20.2	0	0	0	0	0	0	0	32.25
1668410968967227000	0.05	0.1	40.86	0.03	4.31	0	0.48	20.2	0	0	0	0	0	0	0	32.25
1668410967964856000	0.05	0.11	42.66	0.03	4.31	0	0.48	20.2	0	0	0	0	0	0	0	32.25

테이블명	컬럼명	컬럼ID	단위
Data	측정 시간	Time	
	AC 전류값 RMS(미사용)	Irsm1	
	AC 전류값 RMS(미사용)	Irsm2	
	AC 전류값 RMS(사용)	Irsm3	A
	진동센서 raw(미사용)	currvolt1	
	진동센서 raw(사용)	currvolt2	mA
	진동속도1(미사용)	currvolt3	
	진동속도2(사용)	currvolt4	mm/s
	엣지PC 습도값	humidity	

테이블명	컬럼명	컬럼ID	단위
Data	디지털 Input1(임시)	Input1	
	디지털 Input2(임시)	Input2	
	디지털 Input3(임시)	Input3	
	디지털 Input4(임시)	Input4	
	외부온도	temp1	°C
	외부온도	temp2	°C
	엣지PC 온도값	temperature	°C

2.4.2 FANUC 데이터

name : FANUCDATA	Automatic Driving GW DRESSING INTERVAL GW WHEEL REMAIN Macual Automatic Mode Manual Automatic PART COUNTER RW DRESSING INTERVAL RW WHEEL REMAIN Status of automatic operation Status of axis movement TOTAL_COUNTER
166841712964710000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
166841711966370000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
166841710916450000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
1668417091530000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
166841708965380000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
1668417079765495000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
166841706964954000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
166841705952950000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
1668417049492930000 0	0 33.2798 0 0 0 690 9.8867 0 0 0
1668417039627680000 0	0 33.2798 0 0 0 690 9.8867 0 0 0

테이블명	컬럼명	컬럼ID	단위	비고	값 변화
FANUC DATA	측정시간	time	s		값 변화 있음
	생산수량	TOTAL_COUNTER			
	일일 생산수량	PART COUNTER			
	센타레스 GW휠(소재를 깎는 훨) 드레싱 주기	GW DRESSING INTERVAL			
	센타레스 GW휠(소재를 깎는 훨) 잔량	GW WHEEL REMAIN	mm		값 변화 있음
	센타레스 RW휠(소재를 바꿔주는 훨) 드레싱 주기	RW DRESSING INTERVAL			값 변화 있음
	센타레스 RW휠(소재를 바꿔주는 훨) 잔량	RW WHEEL REMAIN	mm		값 변화 있음
	수동/자동	Manual Automatic			
	자동운전중	Automatic Driving			
	AUTOMATIC/MANUAL mode	Macual Automatic Mode	0 : X (정지) 1 : 가동(자동)		값 변화 있음
	Status of automatic operation	Status of automatic operation	0 : 정지(Stop) 1 : 일시정지(Hold) 2 : 가동(Start) 3 : 재가동(Restart)		값 변화 있음
	Status of axis movement	Status of axis movement	0 : X (정지) 1 : 축 동작 (가공) 2 : 축 원위치 (가공) 3 : 축 대기 (가공)		값 변화 있음

2.4.3 MES 데이터

테이블명	테이블 설명
TB_PROD_RSLT	생산 실적
TB_PROD_STOCK	공정 재고
TB_PROD_RSLT_BAD	불량 판정

3. 홈페이지

3.1 제작 홈페이지 상세설명

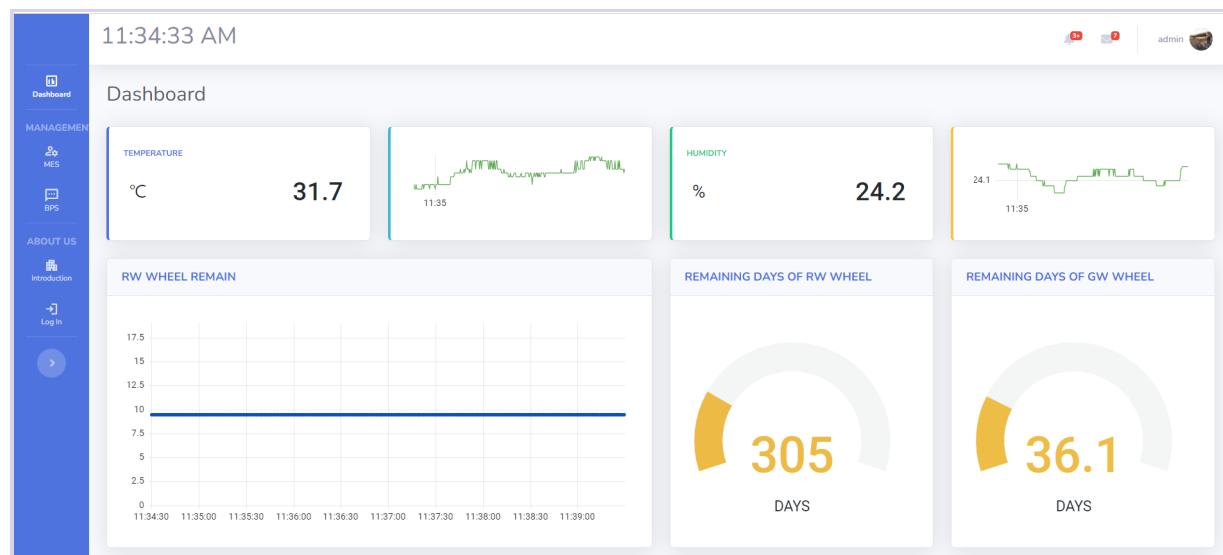
① 메인화면

※ 주요 기능 : 회사의 이미지 공유, 서브 페이지 이동 역할



② 대시보드 화면

※ 주요 기능 : 온도/습도 시스템 상태 , RW 훈 잔량 상태, RW/GW 훈 잔량 예측 잔량에 따른 교체 시기 예측 표시



③ MES 화면

※ 주요 기능 : MES (MES, 생산 관리 시스템은 기업의 생산 현장에서 작업 일정, 작업 지시, 품질 관리, 작업 실적 집계 등 제반 활동을 지원하기 위한 관리 시스템을 말함.)
페이지로서, 생산현황, 재고 개수, 불량품 개수 관련 데이터 표 표시

The screenshot shows the 'Production Management' section of the MES dashboard. On the left sidebar, 'MES' is selected under 'MANAGEMENT'. The main area displays 'PRODUCTION PERFORMANCE' and 'INVENTORY STATUS' tables.

PRODUCTION PERFORMANCE

Time	총 생산 수량
2022-11-25 11:44:00	
2022-11-25 11:45:00	
2022-11-25 11:46:00	
2022-11-25 11:47:00	
2022-11-25 11:48:00	
2022-11-25 11:49:00	

INVENTORY STATUS

Time	입출고 수량
2022-11-25 11:44:00	
2022-11-25 11:45:00	
2022-11-25 11:46:00	224

④ BPS 화면

※ 주요 기능 : 누적 출입고량, 불량품 개수, 기업의 비전과 미션, 장기 목표 리스트, 부서별 요구 사항 입력 폼

The screenshot shows the 'Cooperative Work' section of the BPS dashboard. On the left sidebar, 'BPS' is selected under 'MANAGEMENT'. The main area displays 'Current State', 'Project', and 'Our Mission & Requirements for Each Team' sections.

Current State

This Week Cumulative Receiving Goods Quantity
금주 누적 입고량 : **146**

This Week Cumulative Forwarding Quantity
금주 누적 출고량 : **37.6**

Project

● Our Value
제품주문에서 최종 제품의 완성까지 생산활동의 최적화를 위해 제조현장에서 발생하는 다양한 상황에 유기적으로 신속하게 대응하여 품질 높은 상품을 만들어 고객에게 만족감을 준다.

● Our Mission
뛰어난 사상가와 행동가들의 전문 지식을 최신 컴퓨터 과학 기술과 결합해 모두가 기술을 접하고 사용할 수 있게 한다.

◀ Partner Companies

Our Mission & Requirements for Each Team

We Can Do Whatever Our Customers Want +

- AI 알고리즘 개발 ✕
- Real Time Connection ✕
- Any Where Office 확장 ✕
- 고객사 고정비 및 비교경비 절감 강화 ✕
- Big data 기반의 성능예측 시스템 개발 ✕
- 생산성 개선 30 % ✕
- 협력 회사 주기적인 성능 평가 ✕

Can Your Team Do It For Us? +

- 빅데이터 분석 부서 | 알고리즘 성능 개선 ✕

⑤ INTRODUCTION 화면

※ 주요 기능 : 기업 소개, 팀원 소개, 찾아오는 길, 문의 사항

The screenshot shows the 'Introduction' section of the Acoustic website. On the left, there's a vertical navigation menu with options like MES, BPS, ABOUT US, Introduction (which is currently selected), and Log In. The main content area features a blue header 'CEO 인사말' (CEO Message) with a quote: '"업타운은 정보의 가치와 고객의 미래를 소중히 여깁니다."'. Below the quote is a detailed paragraph about the company's mission and services. A photo of a person wearing a mask and working at a desk with multiple monitors is on the right. At the bottom, there are three numbered sections: 01 고객 맞춤형 (Customer Customized), 02 고급 보호 프로그램 (Advanced Protection Program), and 03 자체 개발 플랫폼 (Self-developed Platform).

⑥ LOGIN 화면

※ 주요 기능 : 로그인, 회원가입, 마이페이지 쟽

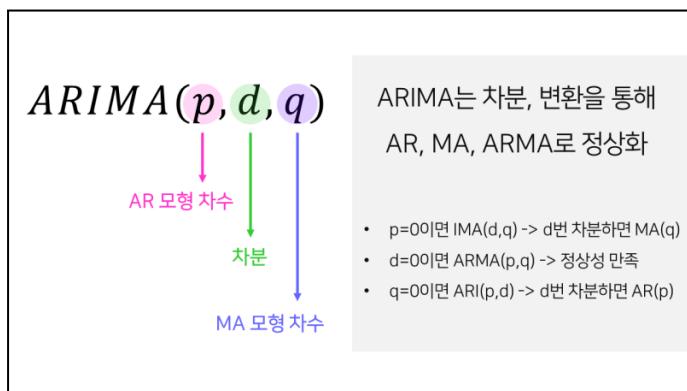
The screenshot shows the login page of the Acoustic website. It has a blue header 'Welcome Back!' and a sub-header '로그인 성공' (Login Success). Below that, it says '"admin" 님 환영합니다' (Welcome "admin"). At the bottom, there are 'logout' and 'Home' buttons.

4. 알고리즘

4.1 사용 알고리즘

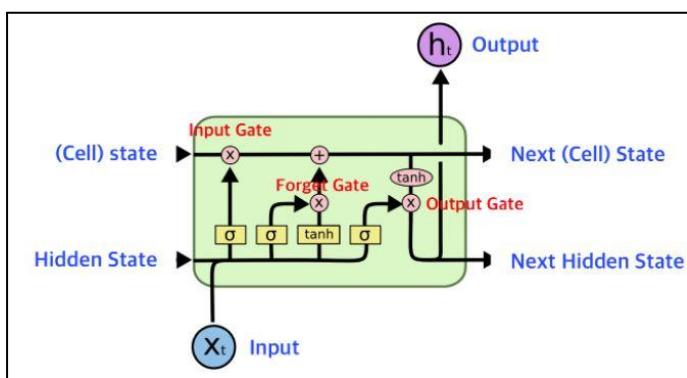
- 월 잔량 예측을 위해 AI 알고리즘을 사용하였으며 잔량을 실시간으로 탐지 후 남은 일수를 예측할 수 있도록 구성하였음.
- 자기회귀와 이동평균을 둘 다 고려하는 시계열 데이터 기반 분석 기법으로 현재 값을 과거 값과 과거 예측값의 오차를 통해 설명하며 예측하는 ARIMA 알고리즘,
- 순환신경망 기법 중 하나로 셀, 입력, 게이트, 망각 게이트를 이용해 기존 순환신경망의 문제점을 방지한 알고리즘인 LSTM 알고리즘을 이용하였음.

4.1.1 ARIMA



- 자기회귀와 이동평균을 모두 고려하여 시계열 데이터 기반 분석 기법으로 현재 값을 과거 값과 과거 예측값의 오차를 통해 설명하며 예측하는 알고리즘

4.1.2 LSTM



- 순환 신경망 기법 중 하나로 셀, 입력 게이트, 출력 게이트, 망각 게이트를 이용해 기존 순환 신경망의 문제점을 방지한 알고리즘

4.2 적용 결과 - PoC(개념실증) 결과 후

- AI 알고리즘을 이용하기에는 각 열의 상관성을 파악해봐도 상관성이 높은 열이 없다.
그러므로 target 값만 가지고 있어도 예측할 수 있는 ARIMA를 제외한 Exponential Smoothing, LSTM 은 사용한다 하더라도 유의미한 결과를 낼 수 없는 것으로 파악하였다.

5. 기대효과

○ 정량적 효과

- AI 데이터 분석을 통해 기존의 월 교체주기 대비 생산량 향상이 10% 이상 늘어날 것으로 기대됨. - 월 잔량 및 교체 주기를 예측하는 AI 분석 모델을 이용하여 이상감지 정확도가 80% 이상 될 것으로 기대됨.

○ 정성적 기대효과

- 설비 상태 및 월 잔량 데이터를 분석하여 설비 예지보전 시스템 구축
- 실시간 모니터링 시스템을 통해 제조 라인에서 작업자들이 실시간 대응이 가능하도록 구현

6. 프로젝트 결과 보고

6.1 방향성 정의

- 월 잔량을 예측할 때 7일을 기준으로 모델을 학습시키고 다음 1일을 예측하는 방식으로 진행했다.
- 교체 시기를 예측할 때 7일을 기준으로 월 감소량을 계산하고 이를 바탕으로 1일 당 교체 시기를 예측했다.

6.2 데이터 전처리

6.2.1 정상성 판별

SARIMAX Results

Dep. Variable: GW_WHEEL_REMAIN **No. Observations:** 107040

Model:	ARIMA(3, 1, 2)	Log Likelihood	741631.755
Date:	Thu, 02 Feb 2023	AIC	-1483251.510
Time:	01:41:00	BIC	-1483194.025
Sample:	0	HQIC	-1483234.118
	- 107040		

Covariance Type: opg

	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	-0	1.18e-25	-0	1.000	-2.31e-25	2.31e-25
ar.L2	-0	1.18e-25	-0	1.000	-2.31e-25	2.31e-25
ar.L3	-0	1.18e-25	-0	1.000	-2.31e-25	2.31e-25
ma.L1	0	1.18e-25	0	1.000	-2.31e-25	2.31e-25
ma.L2	-0	1.18e-25	-0	1.000	-2.31e-25	2.31e-25
sigma2	5.616e-08	4.08e-12	1.37e+04	0.000	5.62e-08	5.62e-08
Ljung-Box (L1) (Q):	0.01			Jarque-Bera (JB):	226309149341.00	
Prob(Q):	0.93			Prob(JB):	0.00	
Heteroskedasticity (H):	1248.41			Skew:	-84.37	
Prob(H) (two-sided):	0.00			Kurtosis:	7124.37	

- Ljung-Box (Q) : 잔차가 백색잡음인지 검정한 통계량

0.930이므로 귀무가설을 기각하지 못한다. 따라서 시계열 모형이 잘 적합되었고 남은 잔차는 더 이상 자기상관을 가지지 않는 백색잡음이다.

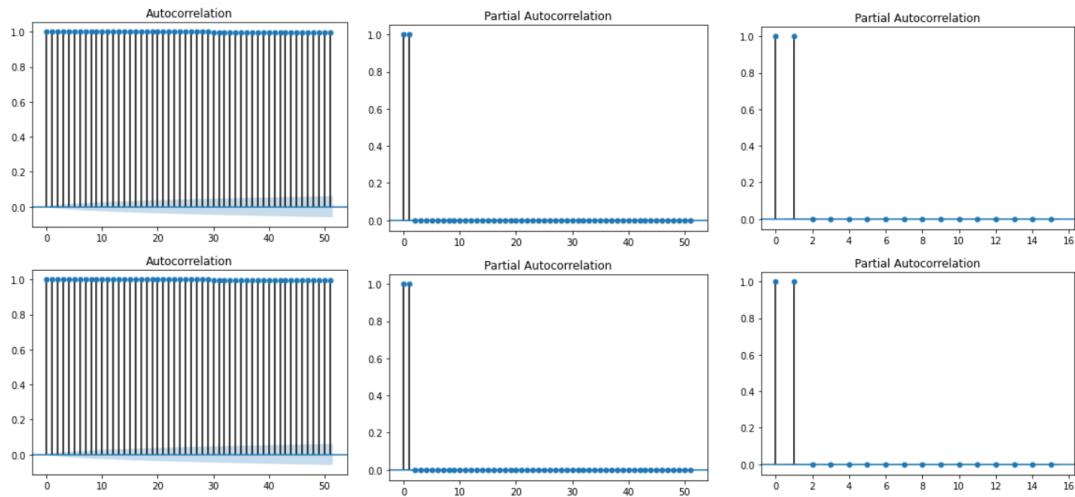
- Heteroskedasticity Prob (H) : 잔차가 백색잡음인지 검정한 통계량

0.000이므로 잔차가 이분산을 띤다.

- Jarque-Bera Prob (JB) : 잔차가 정규성을 만족한다.

0.000이므로 잔차가 정규성을 만족하지 않는다.

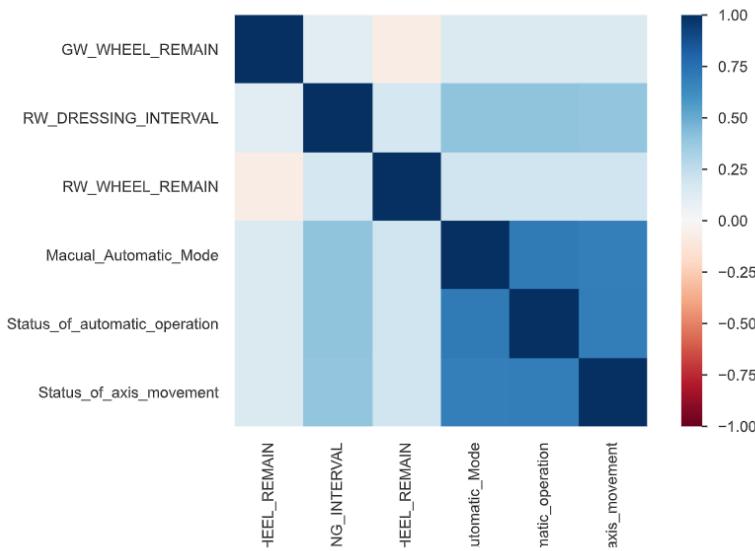
- ACF(자기상관함수), PACF(편자기함수)



데이터에 추세가 있을 때 시계열의 ACF 값은 양의 값을 가진다.

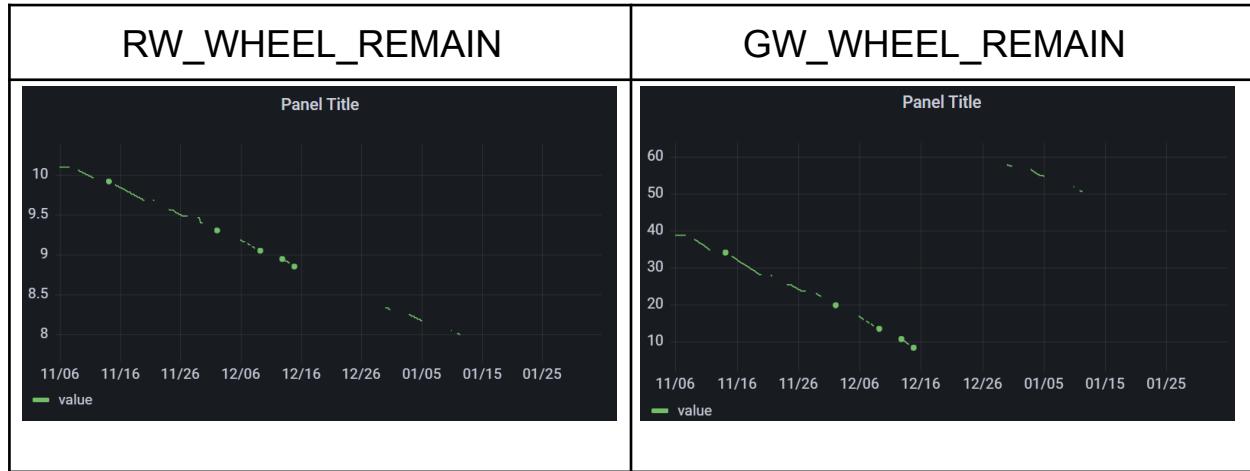
추세가 있고 정상성을 만족하지 않는 데이터라는 결과를 도출했다.

6.2.2 Feature 탐색



예측해야 할 열인 GW_WHEEL_REMAIN과 RW_WHEEL_REMAIN과 상관관계를 보이는 열이 없음을 알 수 있다.

6.2.3 이상치 및 결측치 처리



이상치 처리	결측치 처리
<ul style="list-style-type: none"> - 7일보다 짧은 주기로 훨 잔량의 감소율을 계산하니 예상 교체 날짜가 크게 변경되어서 감소율을 계산하는 기간을 7일로 결정 - 감소율이 평균보다 크게 나타나는 부분은 이상치로 판단 후 이전 값으로 대체 	<ul style="list-style-type: none"> - 'time'열에 결측치 존재함. - 통신 장애로 인해 데이터가 수집되지 않은 날이 많았음. - date_range를 통해 'time'열의 결측값을 채워줌. - 'time'열의 새로운 행이 생기면서 다른 열의 결측치가 새로 생기게 됨. - 교체된 날이 기록되지 않아서 이전, 이후 값을 토대로 전체 기울기를 계산하여 교체 지점을 예측하여 지정 - 'RW_WHEEL_REMAIN'과 'GW_WHEEL_REMAIN'열의 결측치를 선형 보간법을 통해 보간

6.3 최종 모델링

<최종 사용 데이터 정보>

- 데이터 총량 : 91,364,209 개

6.3.1 최종 모델링 시 제외한 feature

Data

- AC 전류값 RMS 1, 2
 >>> 변화량이 거의 없으므로 제외
- 진동센서 1, 3
 >>> 값이 모두 동일하여 제외
- 디지털 Input 1 ~ 4
 >>> 값이 모두 결측치이므로 제외
- 외부온도 1, 2
 >>> 외부온도에 변화가 거의 없으므로 영향이 없다고 판단하여 제외

fanuc

- 생산수량, 일일 생산수량
 >>> 값이 모두 결측치라 제외
- 센타레스 GW휠 드레싱 주기
 >>> 값이 모두 결측치라 제외
- 수동 / 자동, 자동운전종
 >>> 값의 변화가 없어 제외

TB_PROD_RSLT, TB_PROD_RSLT_BAD, TB_PROD_RSLT_STOCK

- 공장구분, 라인 / 설비 코드, 작업지시번호, ...
 >>> QTY로 끝나는 열이름을 제외한 대부분의 열이 불량품 또는 훨의 두께에 영향을 주지 않는 열이라서 모두 제외

6.3.2 함수 계산식



- 노란 그래프 : 전처리 전, 녹색 그래프 : 전처리 후 (계산을 통해 보간한 그래프)

- 1) 시계열 데이터를 받았을 때 빈 시간대가 너무 많아서 데이터를 사용할 수 없는 상황이었다. (빈 시간을 다 채울 경우 빈 행이 70% 이상)
- 2) 심지어 훨 교체주기를 학습해야하는데 교체하는 시점의 데이터가 오류로 인해서 전부 사라진 상태였다. (사라진 구간 : 2022년 12월 14일 오후 ~ 12월 30일 오전)
- 3) 훨을 교체했을 때의 초기두께를 알아낸 후 전체 그래프에서 데이터가 사라지기 전까지의 데이터의 기울기를 얻어냈다.
- 4) 2022-11-04 (39.62mm) ~ 2022-12-14 (8.44mm) (총 40일)
- 5) $\frac{y_2 - y_1}{x_2 - x_1} = \text{기울기}$

$$\frac{8.44 - 39.62}{40} = -0.78$$
, 기울기는 -0.78 이다.
 따라서, 하루 당 감소량이 0.78이라는 결론을 도출했다.
- 6) 구한 기울기를 토대로 빈 구간을 채우려고 계산하니 훨의 두께가 0이 될 때까지 사용해야 빈 구간을 채울 수 있다는 결론이 도출되었다.
- 7) 그 결론에서 얻을 수 있었던 것은 훨 두께가 0이 될 때까지 사용할 시 불량품의 개수가 증가 할 수 있다는 것이었다.
- 8) 훨의 두께가 얕아질 때 불량품 개수가 많아진다면, 훨의 교체시기를 예측 가능하다 판단하였다.
- 9) 하지만 불량품 개수를 시간대 별 그래프로 나타내었을 때, 훨의 두께는 불량품 개수에 영향을 주지 않았다. 작업량의 증가가 불량품의 개수에 영향을 주었다.
- 10) 훨의 두께가 0mm가 될 때까지 사용했다는 가정하에, 훨의 교체시점을 기울기(위의 3번 설명 참고)를 참고하여 예측한 뒤, 그 시점에 훨의 초기두께 60mm를 입력하였다.
- 11) 교체 시점을 정한 후, 보간법을 사용하면 결측치가 사라진 완전한 그래프를 얻을 수 있다.
- 12) 상관성이 없어보이는 열들을 제외하고 LSTM을 사용하려 하니 독립변수가 남아있지 않음
- 13) ARIMA는 Target값만 있어도 예측가능하기에 ARIMA를 사용
- 14) ARIMA도 예측값에 이상이 있음
- 15) 정상성을 확인해보니 정상성을 만족안함(시계열 알고리즘 사용 불가능)
- 16) 그래프가 꾸준한 우향성을 그리고 있으니 하루당 감소량을 이용하여 교체시기를 예측할 수 있을거라 판단
- 17) 완전한 그래프에 구해놓은 전체 기울기를 적용하니 (-0.78) 기울기가 맞지 않는 것을 확인
- 18) 기울기를 전체기간에서 구한것이 큰 오차를 발생시켰다고 판단
- 19) 기울기를 현재 시점에서 일주일 뒤까지만 들고와서 계산하도록 변경
- 20) 변경한 기울기로 계산하자 교체시점과 거의 동일한 남은 일수를 확인
- 21) 일주일 뒤까지 들고오게 변경하자 훨 교체시점에 급격한 기울기 증가 발생
- 22) 기울기가 양수가 되면서 교체일수 예측에 문제가 발생 (남은 일수 : - 777)
- 23) 기울기가 양수가 될 시 전의 기울기를 가져와서 덮어쓰게 변경
- 24) 시계열 알고리즘 특성상 긴 기간을 예측할수록 정확도가 급감하기 때문에 오히려 정확할 수 있다고 파악됨.

6.3.3 LSTM

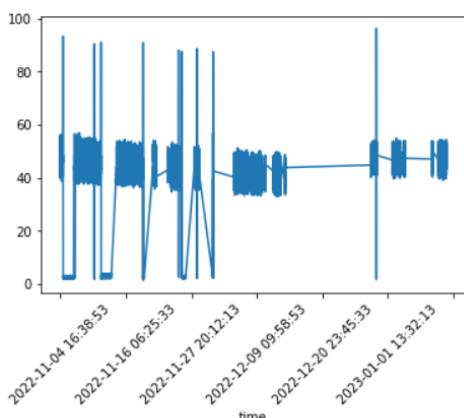
앞서 '6.2.2 Feature 탐색'에서 확인하였듯이 예측해야 할 열인 'RW_WHEEL_REMAIN, GW_WHEEL_REMAIN'과 상관관계를 띠는 열이 존재하지 않는다. 하지만, LSTM을 진행하기 위해 독립 변수가 필요하므로 상황상 관련성이 높을 것으로 예상되는 3가지 feature 'Irsm3, currvolt4, BAD_QTY'들을 이용하여 모델링을 진행하였다.

< 사용한 열 >

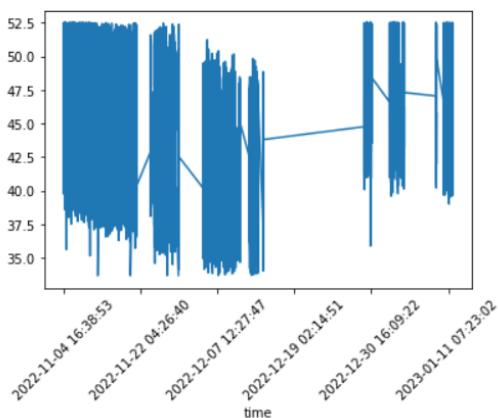
- 독립 변수 X : Irsm3, currvolt4, BAD_QTY
- 종속 변수 Y : RW_WHEEL_REMAIN, GW_WHEEL_REMAIN

< 이상치 처리 >

- Irsm3

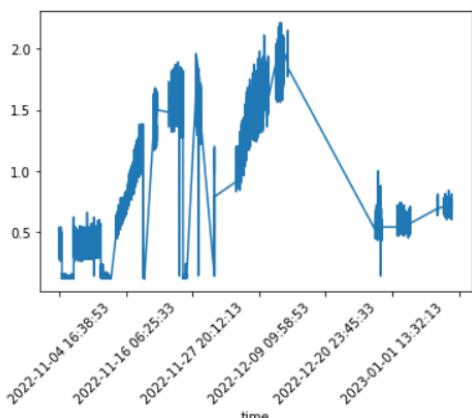


이상치 처리 전

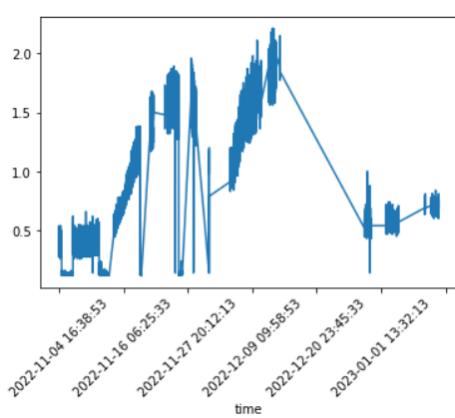


이상치 처리 후

- currvolt 4



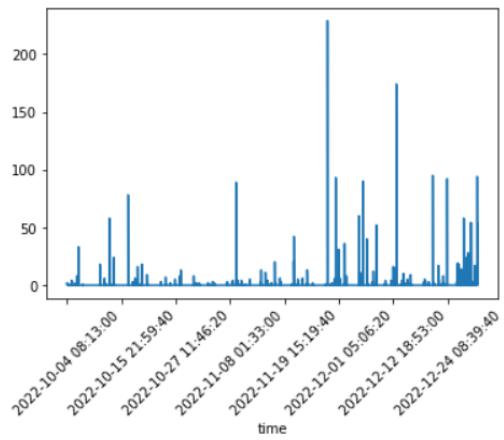
이상치 처리 전



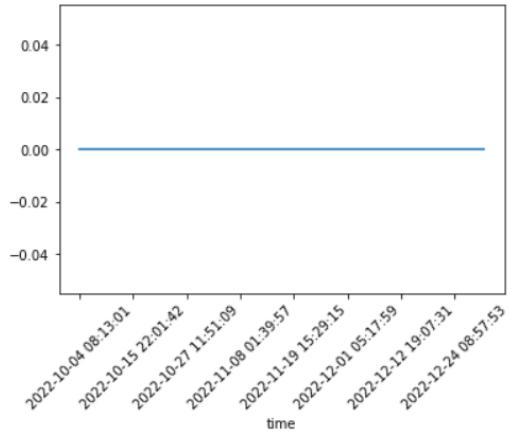
이상치 처리 후

→ 큰 차이가 없다.

- BAD_QTY



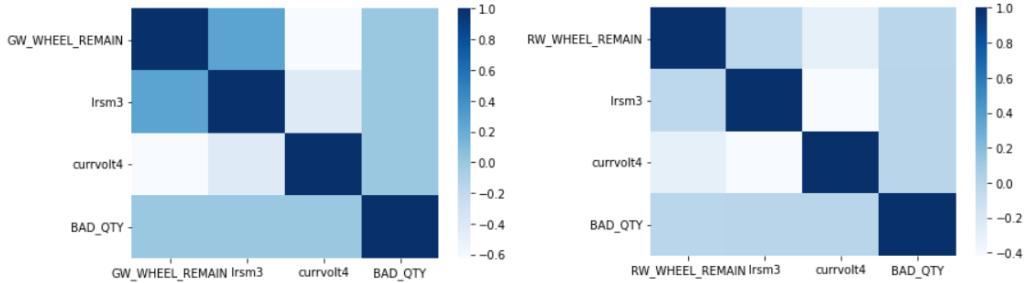
이상치 처리 전



이상치 처리 후

→ **BAD_QTY**는 불량 발생 제품의 개수이므로 열 특성상 이상치를 제거하면 안 될 것으로 판단된다.

< 상관 분석 >



→ **GW_WHEEL_REMAIN**과 **Irsm3** 둘 사이의 상관관계가 약 0.4인 것을 제외하면 **Feature**들 간의 상관관계는 거의 없다.

< 모델링 진행 >

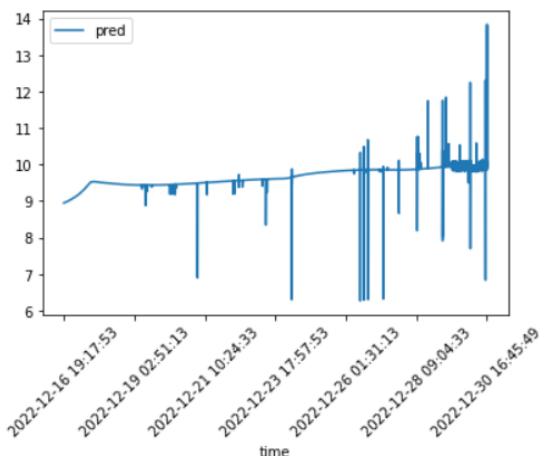
train 데이터와 **test** 데이터는 7:3의 비율로 나누었다. **train** 데이터를 이용해 모델을 학습시키고 **test**데이터를 통해 예측을 진행하였다.

모델링은 Keras의 Sequential 모델을 이용하였다. Activation함수는 시그모이드 함수의 기울기 소실 문제를 해결한 relu를 이용하였고 optimizer는 adam, 손실 함수는 mse를 사용하였다.

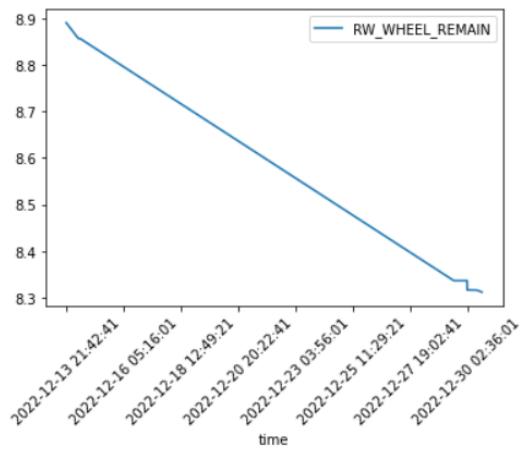
test 데이터를 이용해 예측을 진행한 후 얻어진 y값과 본래의 y 값의 분포를 시각화하여 결과를 비교하였다.

< 결과 >

● RW_WHEEL_REMAIN

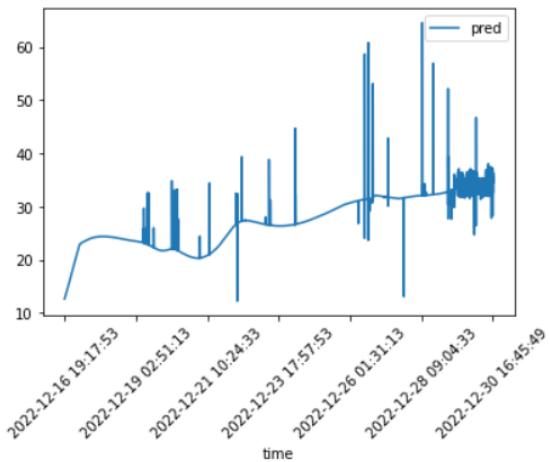


예측한 그래프

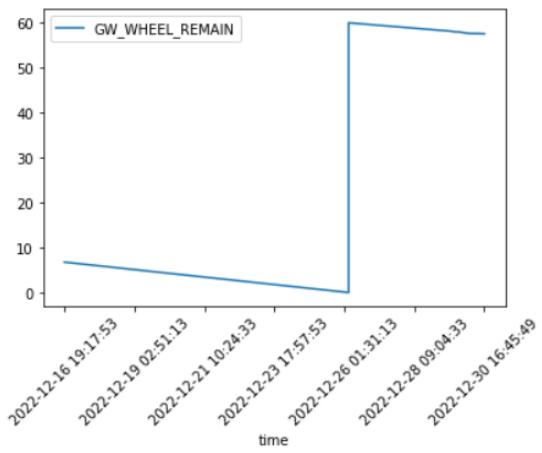


검증용 그래프

● GW_WHEEL_REMAIN



예측한 그래프



검증용 그래프

LSTM을 사용하고자 하였으나 독립변수들간의 상관성이 낮았기 때문에 예측이 제대로 되지 않은 것을 알 수 있었다.

6.3.4 ARIMA

ARIMA는 하나의 target값을 이용해 시계열 분석을 진행할 수 있는 알고리즘이다.
ARIMA의 차수 p,d,q 값을 얻기 위해 **auto arima**를 실행해서 **best model**을 구했다.

```
Performing stepwise search to minimize aic
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=-10250185.215, Time=212.11 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-10223672.280, Time=82.45 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-10245366.942, Time=104.53 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=-10241692.866, Time=324.58 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=-10223143.811, Time=25.26 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=-10250114.441, Time=190.16 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=-10250054.021, Time=176.38 sec
ARIMA(3,1,2)(0,0,0)[0] intercept : AIC=-10250204.681, Time=237.42 sec
ARIMA(3,1,1)(0,0,0)[0] intercept : AIC=-10250110.039, Time=181.26 sec
ARIMA(4,1,2)(0,0,0)[0] intercept : AIC=-10250193.457, Time=302.96 sec
ARIMA(3,1,3)(0,0,0)[0] intercept : AIC=-10250202.716, Time=301.47 sec
ARIMA(2,1,3)(0,0,0)[0] intercept : AIC=-10250204.575, Time=273.29 sec
ARIMA(4,1,1)(0,0,0)[0] intercept : AIC=-10250173.053, Time=266.81 sec
ARIMA(4,1,3)(0,0,0)[0] intercept : AIC=-10250200.567, Time=332.71 sec
ARIMA(3,1,2)(0,0,0)[0] : AIC=-10250078.442, Time=314.53 sec

Best model: ARIMA(3,1,2)(0,0,0)[0] intercept
Total fit time: 3325.943 seconds
```

⇒ 최적 모델인 ARIMA(3,1,2)를 이용해서 모델링을 진행했다.

정확도를 높이기 위해 처음 7일을 이용해 모델을 **fit** 한 후, 다음 1일을 **forecast**하는 방식을 이용했다.

RW 훈련량에 대한 MAE 값은 0.012, GW 훈련량에 대한 MAE 값은 0.297을 기록했다. GW 훈련량의 MAPE값은 3.792로 기록되었는데 이는 결과값에 0에 가까운 값들이 많이 포함되었기에 이러한 결과가 나타났다.

제공받은 RW 훈련량 데이터에 교체된 구간에 대한 데이터가 누락되었기에 그 부분에 대한 데이터가 보완된다면 조금 더 나은 결과를 도출할 수 있을 것으로 기대된다.



- 노란 그래프 : GW 월 잔량, 녹색 그래프 : arima를 통해 예측한 GW 월 잔량



- 노란 그래프 : RW 월 잔량, 녹색 그래프 : arima를 통해 예측한 RW 월 잔량

● 모형 평가

	RW 월 잔량	GW 월 잔량
MAE	0.012	0.297
RMSE	0.018	0.406
MAPE	0.129	3.792
R^2	0.999	0.998

6.4 DB 설계

DB설계 (테이블 개)

- DB 이름 : SensorData
- measurement 이름 : Value

이름	내용	데이터 형식	단위
time	측정 시간	timestamp	
Irsm3	AC RMS 전류값	String	A
currvolt2	진동센서 raw값	String	mA
currvolt4	진동 속도	String	mm/s
humidity	엣지PC 습도 값	String	%
temperature	엣지PC 온도 값	String	°C
GW_WHEEL_RemAIN	GW휠(소재를 깎는 휠) 잔량	Float	mm
RW_WHEEL_RemAIN	RW휠(소재를 깎는 휠) 잔량	Float	mm
Macual_Automatic_Mode	자동 / 수동 모드		
Status_of_automatic_operation	자동 작동 단계	int	0 : 정지 1 : 일시정지 2 : 가동 3 : 재가동
Status_of_axis_movement	축 동작 상태		0 : 정지 1 : 축동작 2 : 축 원위치 3 : 축 대기

RW_DRESSING_INTERVAL	RW월 드레싱 주기	float	
BAR_QTY	바코드 발행 수량	int	
GOOD_QTY	양품 수량	int	
TOT_QTY	총 실적 수량	int	
BAD_QTY	불량수량(폐기)	int	
IN_QTY	최초 입고 수량	int	
STOCK_QTY	재고 수량	int	
GW_Remain_Day	GW월 남은 일수 예측	float	
RW_Remain_Day	RW월 남은 일수예측	float	

7. 개선점 및 발전 사항

7.1 개선점

개선점	내용
데이터 부족	데이터의 대부분이 훨씬 두께와 상관없는 열밖에 없었다. 따라서 연관성이 있는 데이터가 있었다면 다양한 알고리즘을 사용해서 더 좋은 결과가 나왔을 것.

7.2 발전 사항

발전 사항	내용
실시간 연동	기업 특성상 실시간 데이터를 제공해 줄 수 없다고 해서 이전 시점의 데이터만 사용해서 예측을 하였다. 하지만 데이터를 실시간으로 받을 수 있다면 DB에 실시간으로 저장하고 바로 전처리 후 모델에까지 적용 할 수 있을 것.

8. 프로젝트 소감

강보원

처음에 현장에서 실제로 수집되는 데이터를 사용할 수 있다는 생각에 기대를 안고 프로젝트를 시작했다. 하지만, 가공되어 있는 데이터만을 이용해서 분석을 진행하다 전혀 가공되지 않은 데이터를 접했을 땐 어떻게 시작해야 할지 막막했다. 또한, 데이터의 방대한 양 때문에 하나의 모델을 가동하는 데에도 시간이 오래 걸렸다. 그렇지만 시계열 데이터의 전처리, 모델링에 대해 많이 학습할 수 있는 기회가 되었다고 생각한다. 이번에는 2개의 모델만 사용해보았지만 다음에는 더 다양한 모델과 다양한 데이터를 이용한 시계열 분석에 도전해보고 싶다.

박현명

‘실제 실무 현장 데이터를 다루는 주제이니 취업 후에 많은 도움이 되겠다.’라는 생각으로 해당 프로젝트를 선택했다. 하지만 실제 현장 데이터를 마주해보니 살벌했다. 이상을 바라보는 기획자와 현실을 직면하고 있는 개발자가 서로 다른 방향을 보고 있다는 어두운 진실을 알게 되었다. 이번 프로젝트를 통해 데이터분석의 현실을 맛보았으며, 앞으로 마주하게 될 살벌한 데이터들을 미리 체험할 수 있어서 이번 프로젝트는 의미가 깊었다. 마음의 준비를 하고, 현장 데이터를 잘 다룰 수 있는 분석가가 되어야겠다.

최정인

분석한 데이터를 효율적으로 고객사에게 보여줄 수 있는 방법에 대해 고민하였고, 이를 통해서 다양한 서비스 또한 기획할 수 있는 시간을 얻었다. 또한, 힘들게 고민하고 해결한 결과물을 시각적으로 전달하는 것이 얼마나 중요한지 알게 된 계기가 되었다. 데이터 분석팀들의 노고가 잘 비추어졌길 바라며 이만 소감을 마치겠다...

최창규

실무데이터가 기대한것보다 훨씬 복잡해서 사용하는데 애먹었지만 그 덕분에

많이 배웠다.