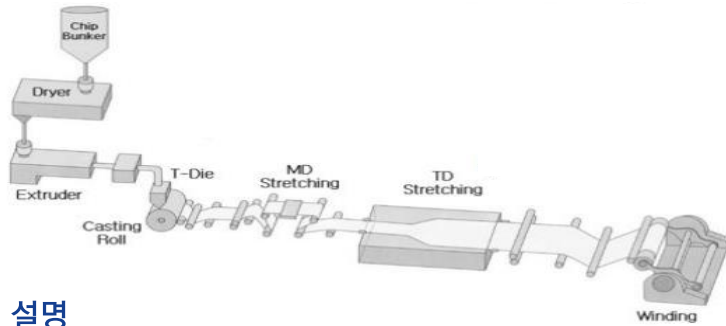


LS 빅데이터 스쿨 3기 추가실습



1. 개요 및 데이터 설명



TORAY 중합 공정라인
매 분마다 400여 개의 변수 발생

[문제] 400여 개의 변수 중 급감 또는 급상승 현상을 보이는 이상 변수 발생

데이터 설명

원 데이터 파악

- ✓ 중합 공정라인(CPS-8)에서 수집한 데이터
- ✓ 데이터 수집 단위: 1분
- ✓ 2가지 이상 데이터로 구분
 - 1. 이상 데이터 ▶ 2018년 12월 8일 색조 L치(LAB_8CHIP_L) 변수의 이상을 포함한 한 달 데이터 (중합 CPS-8 DB B1, 중합 CPS-8 DB B2)
 - 2. 이상 데이터 ▶ 2019년 4월 8일 방사C 펌프RPM(8SIC_P4360C.PV) 변수의 이상을 포함한 한 달 데이터 (중합 CPS-8 DB D1, 중합 CPS-8 DB D2)

시점 불일치 처리 필요

- ✓ CPS_8_DB_B1, CPS_8_DB_B2 데이터의 Timestamp 시점 불일치
→ '중합 CPS-8 DB B1' 데이터의 2018-11-09일자 데이터 삭제 필요

데이터 중복 처리 필요

- ✓ 데이터 병합 시, 중복되는 컬럼 및 행 삭제 필요

전처리 전							전처리 후			
구분	데이터명	행	변수	시점 불일치	중복 변수	중복 행	Timestamp		행	변수
이상 데이터 1	CPS_8_DB_B1	44,641	199	○	○	○	2018-11-10 00:00:00	2018-12-10 00:00:00	43,201	410
	CPS_8_DB_B2	43,201	213	○						
이상 데이터 2	CPS_8_DB_D1	44,641	199		○		2019-03-09 00:00:01	2019-04-09 00:00:01	44,641	410
	CPS_8_DB_D2	44,641	213							

2. 목표 및 이상 데이터

프로젝트 목표

특정 변수 이상원인탐지

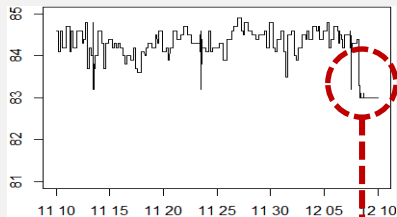
특정 변수의 이상에 영향을 미친 과거 시점의 원인 변수를 찾고, 변수 간 인과관계를 통계적으로 검정

이상 데이터

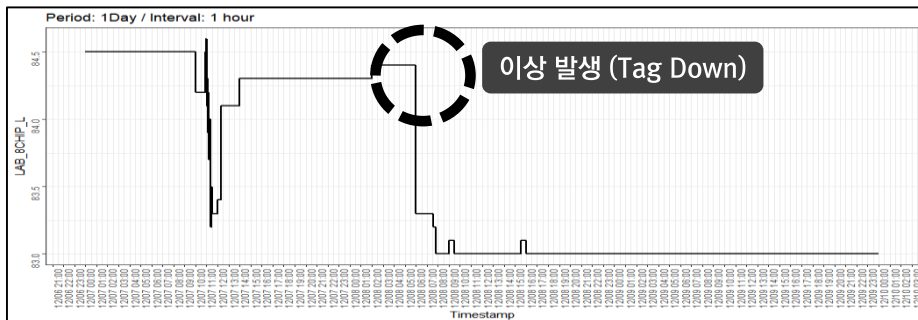
이상 데이터 (1)

색조 L치(LAB_8CHIP_L) 이상원인탐지 모델

● 이상변수 특징



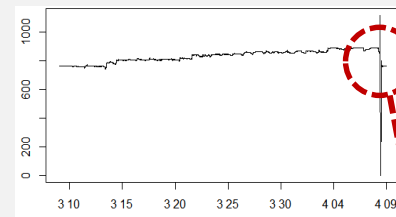
- ✓ 이상 전 : 81~85사이의 값을 가지되 일정 주기(10일 이내)로 등락 반복
- ✓ 이상시점 : 12월 8일 값 급감
- ✓ 이상 후 : 급감한 83 부근에서 값 유지



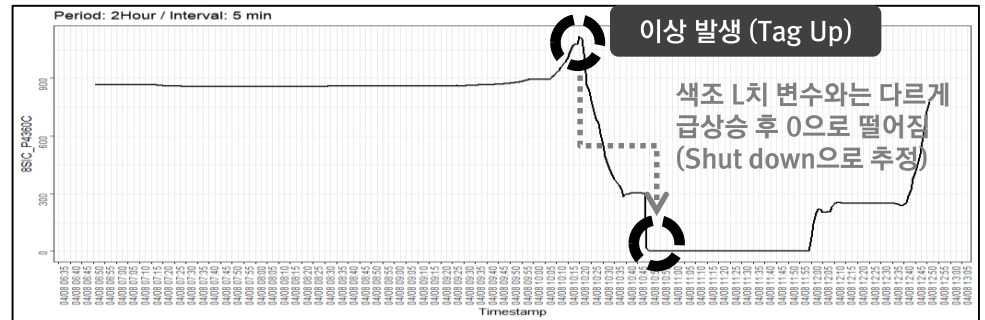
이상 데이터 (2)

방사C 펌프RPM(8SIC_P4360C.PV) 이상원인탐지 모델

● 이상변수 특징



- ✓ 이상 전 : 741~895사이의 값을 가지되 이상시점에 가까질수록 우상향 추세
- ✓ 이상시점 : 4월 8일 값 급상승
- ✓ 이상 후 : 급상승 후 0까지 떨어졌으나 다시 800부근으로 돌아옴

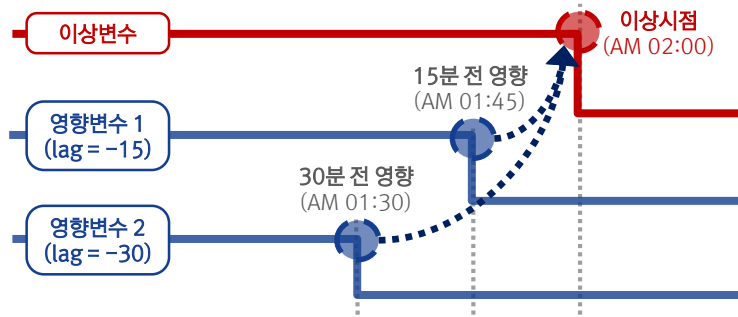


3. 교차상관 예시

➔ 교차상관 분석을 실시하여 이상 데이터와 시계열적 상관도가 높은 영향변수를 선별

영향변수와 영향시점(Lag)

- 영향변수 : 이상변수와 시계열적 상관도가 높은 변수
- 영향시점 : 시계열적 상관도를 갖는 시점을 의미하며 단위는 '분'



최대영향시점

- 정의 : 영향관계를 검토할 과거 시점의 최대치
- 가정 : 최대 이틀 전(maxlag = -2880)까지의 영향관계가 의미 있을 것이다



교차상관 결과

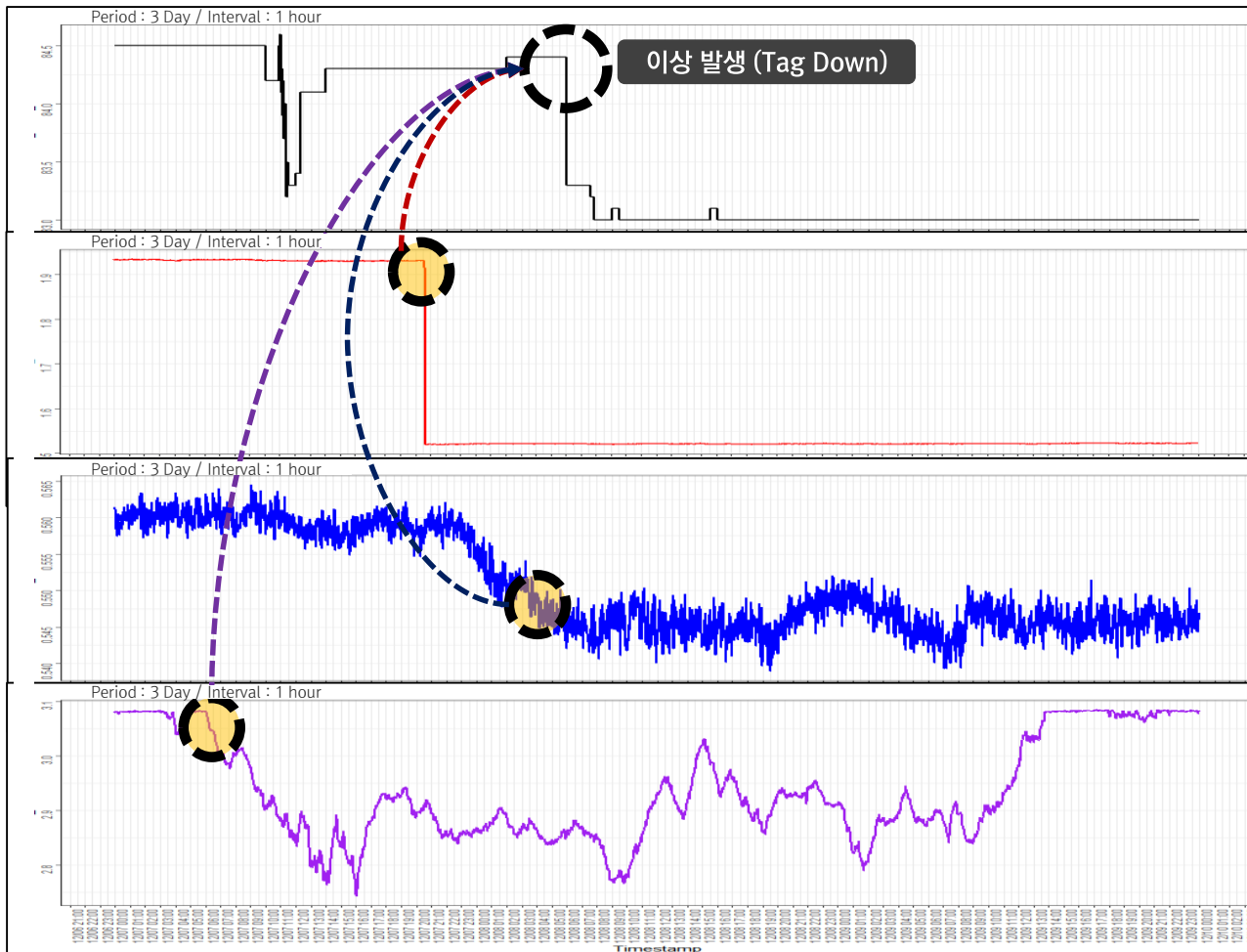
- **이상변수와 나머지 변수**는 매 시점마다 0~1사이의 영향도를 가짐
- 두 변수 간 영향도를 최대영향시점(maxlag)까지 산출한 후 가장 큰 값을 갖는 영향도를 '최대영향도'로 간주하고 그때의 '영향시점'을 결과로 정리
- 최대영향도의 경우 통계적인 관점에서는 0.6 이상의 값을 가져야 영향도가 높은 변수로 판단 가능
- 영향시점이 0이다 = 이상변수와 영향정도가 최대가 되는 시점이 과거가 아님

번호	변수명	최대영향도 (교차상관계수)	영향시점 (단위 : 분)
1	이상변수	1.000	0
2	변수1	0.69	-188
3	변수2	0.67	-10
4	변수3	0.65	-344
5	변수4	0.62	0
6	변수5	0.60	-77
7	변수6	0.55	0
...

이상변수

영향변수 Top 3

3. 교차상관 예시



이상변수

변수명 : 이상변수
 이상시점 : 12.08.오전 04:00 경
 이상현상 : Tag Down. 이후 83
 부근 지속

영향변수 Top 3

변수명 : 변수1
 최대영향도 : 0.69
 영향시점 : 188분 전

변수명 : 변수2
 최대영향도 : 0.67
 영향시점 : 10분 전

변수명 : 변수3
 최대영향도 : 0.65
 영향시점 : 344분 전