

미래원 DT 교육용 샘플데이터 설명

2018. 01



1. 데이터 상황 설명

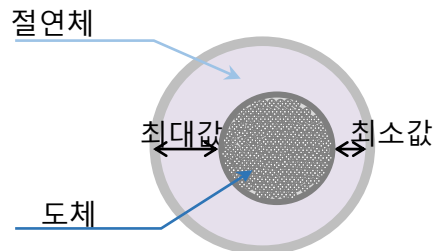
LS전선에서는 CCV절연공정을 통해 Cable을 생산한다. 해당 공정은 원재료인 도체와 절연재를 투입 하여 도체에 절연체를 입히는 공정으로 절연체는 방전 및 외부 요인으로부터 도체를 보호하는 역할을 한다. 구체적인 Cable 생산 과정은 우선 도체가 투입되면서 시작된다. 투입된 도체는 메터링을 거쳐 압출기에서 녹여져 나오는 절연수지를 두르게 된다. 적절한 두께로 둘러싼 절연수지는 고온, 고압의 환경인 가교관을 지나면서 굳혀지기 시작하는데 처음 유동성이 있던 수지는 가교관 끝 쪽에서 굳혀진 상태로 출력이 된다. 위 과정을 거쳐 Cable의 형태가 완성되면 냉각관에서 생산 중 발생한 열을 식히고, 작업자가 일정 길이(조장)로 잘라 드럼에 감아 제품을 완성시킨다.

한편, 생산 과정 중 Cable은 편심불량이 발생할 수가 있다. 편심 불량이란 절연체가 일정한 두께로 도체를 감싸지 않아 Cable의 단면을 확인했을 때, 도체가 중앙에서 어긋나 있는 것을 의미한다. 그 원인으로는 편심 불량은 절연수지가 아직 굳지 않은 압출기와 가교관 사이에서 주로 발생한다고 추정된다. 이러한 편심불량의 정도는 편심율을 통해 계산되며, LS전선에서는 편심율이 30%이상인 경우 불량으로 판정하고 있다. LS전선에서는 편심율을 낮추기 위해 여러 노력을 하고 있으며, 19년도 편심율을 30%에서 15%로 감소시킬 계획이다. 따라서 빅데이터 분석이 편심율을 감소시킬 수 있을지 알아보려고 한다.

편심불량이란?

전선의 도체가 절연체 중심점에서 치우친 정도

: 편심율 (%) = (최대값 - 최소값) / 최대값



CCV 절연공정



LS전선 공정을 바탕으로 생성한 가상 데이터

항목 명칭	변수 구분	데이터 타입
Data Time	-	Timestamp
편심률_계측기1	Y1	Numeric
편심률_계측기2	Y2	Numeric
도체 예열 온도	X1	Numeric
압출기 수지온도	X2	Numeric
압출기 모터RPM	X3	Numeric
압출기 압력	X4	Numeric
압출기 실린더 온도	X5	Numeric
냉각기 입구 온도	X6	Numeric
냉각기 출구 온도	X7	Numeric
가교관 온도	X8	Numeric
조장 길이	X9	Numeric

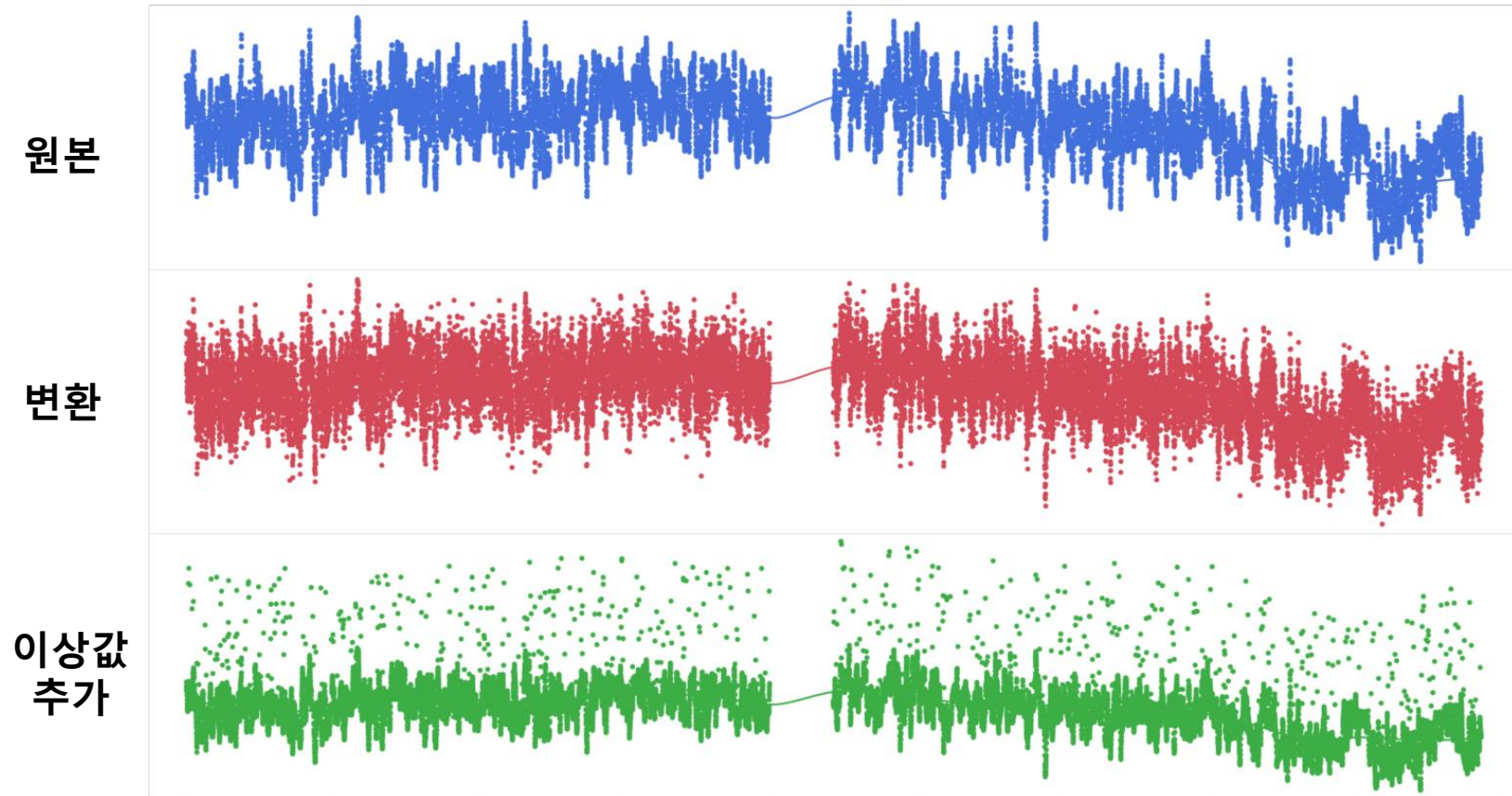
Time series 데이터

- 목표변수(y) : 2개
- 독립변수(x) : 9개

데이터 보안성 향상안

- 실제 데이터를 기반으로 각 항목별 실데이터 * 난수를 곱함 (정보유출 방지)
- 일부 항목(y2)은 분석 교육 목적에 맞게 가공하였음
- 항목 명칭을 분석 교육에 적합하게 간소화함

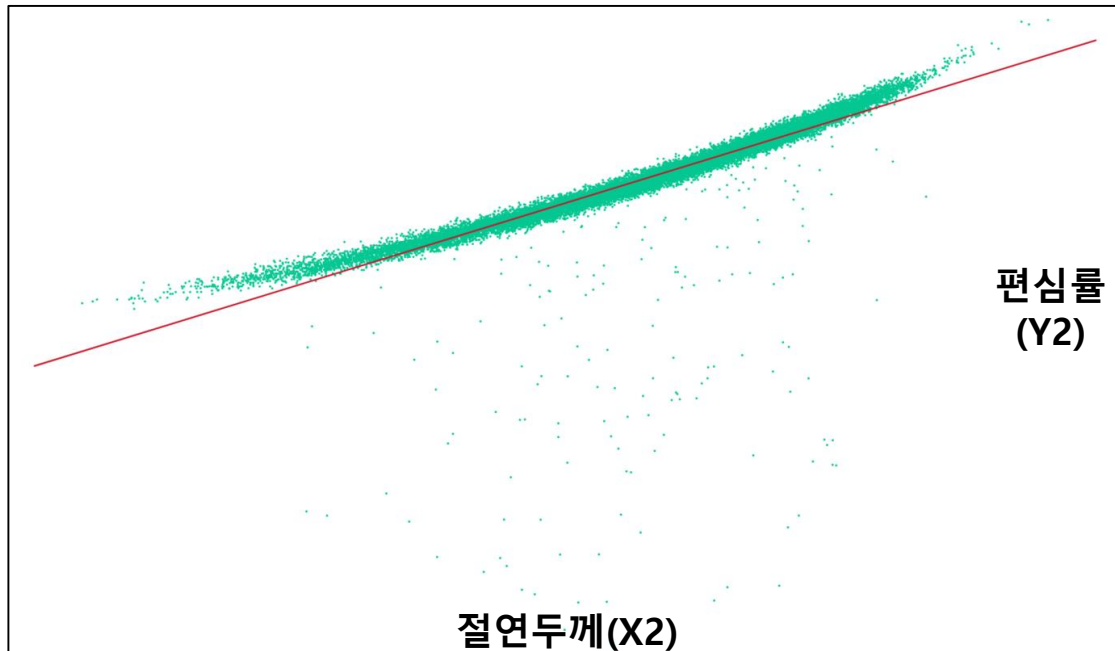
분포에서 Random한 값을 추출하여 데이터를 변환시켜 원본 데이터 값 변형함



데이터 값 변환식 | $\frac{x + \text{norm}(x, \alpha)}{2}$

이상 값 추가식 | $x + \text{norm}(0, \alpha) \cdot \text{unif}(0, 10) \cdot \text{int}(0, 1) \cdot \text{int}(-1, 0)$

데이터 분석 결과 예시



적용 가능한 분석 방법 (예)

1. 데이터 형식 변경
2. 이상값 제거
3. 결측값 제거/보간
4. 기술통계
5. 회귀 분석
6. 로그 그래프 변경

그 외 교육 시 필요에 따라 추가 적용 가능

데이터 분석 결론 | 편심률은 공정 후반으로 갈수록 높아지며, 압출기 수지압력과 높은 상관관계를 갖는다

* 해당 내용 및 결론은 교육용 예시이므로 실제 공정 및 분석과 많은 차이가 있음

실제 분석과정에서 고려해야 할 차이

- 실제 데이터 분석에서는 더 많은 분석 기법들과 전처리 과정을 요구함
- 단일 요인이 목표변수에 영향을 미치는 것이 아니라 다양한 요인이 영향을 미침
- 시간에 따라 동일한 공정 조건이라도 결론이 바뀔 수 있음 (시계열성)
- 도메인 지식과 통계 지식을 동시에 요구하므로 장기간 협업을 통해 발전 시켜야 함