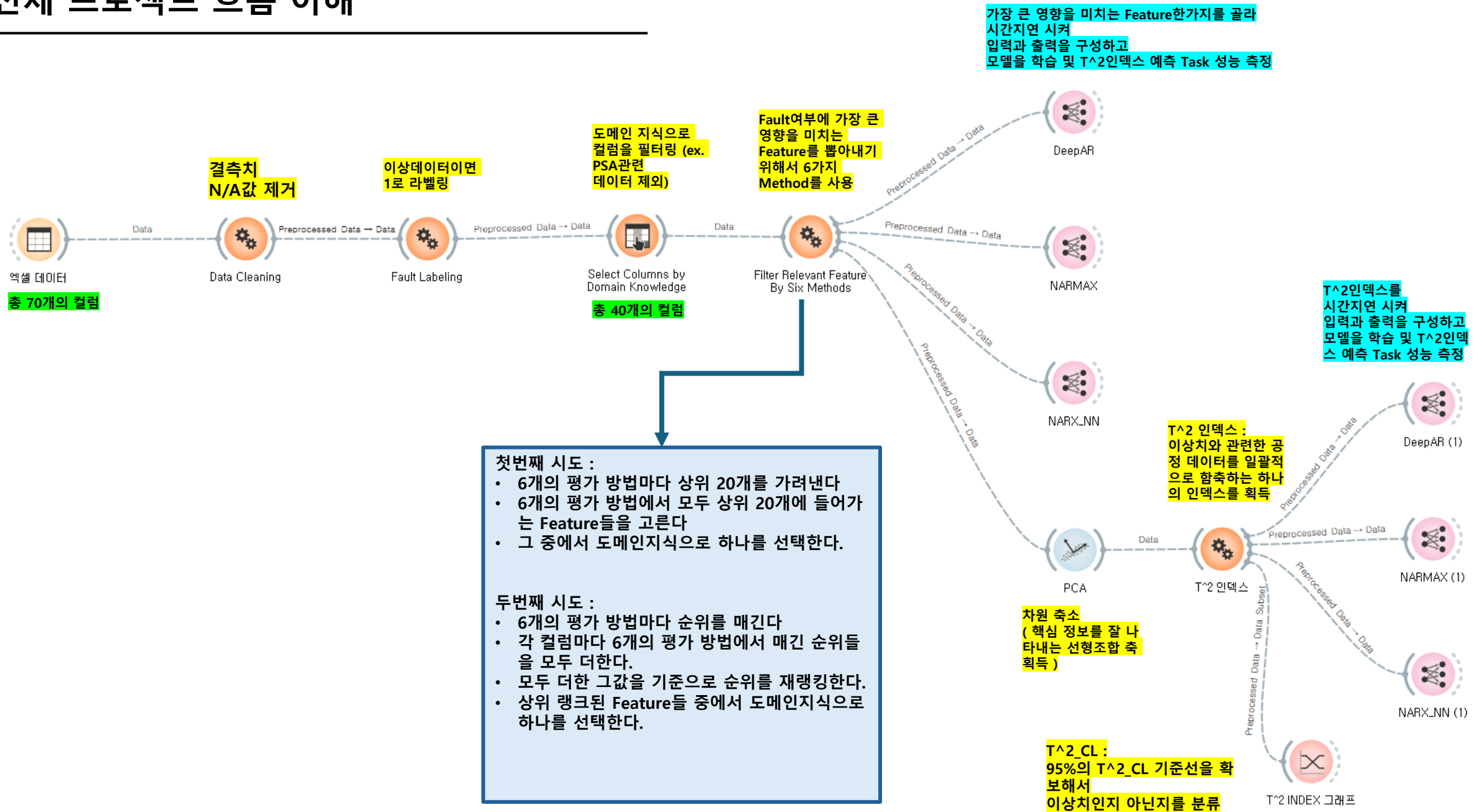


장교진 박사님께 SMR 이상감지 프로젝트 관련 질의응답 요청

안용상

Date : 2024 / 07 / 09

전체 프로젝트 흐름 이해



Question 01

이전 슬라이드에서
박사님께 프로젝트 설명을 들은 정보와
코드 구현 내용을 종합해서
프로젝트 전반적인 흐름을 이해한 것을 다이어그램으로 표시했습니다.

제가 이해한 부분 중에 혹시 부족하거나 오류가 있는 부분이 있는지 한번 봐주시길
부탁드리고 싶습니다!

Question 02

제공해주신 1차데이터 (660데이터 포인트)에 대한 플롯입니다.

주피터 파일을 보니, 박사님께서 100번째 데이터 포인트부터 Fault라고 라벨링을 하신 것을 확인할 수 있었는데요,

100번째부터 Fault라벨링을 했다고 한다면,
우측 사진의 **빨간 박스**안에서 요동치기 시작하는 부분부터 Fault가 라벨링 됩니다.

그러나 조금 더 뒤쪽을 확인해보면 더 급격한 변화가 감지되는 부분(**노란박스**)이 있습니다. 일관적으로 360번째 데이터부터 큰 변화가 감지되었습니다.

일관적으로 더 급격한 변화를 나타내는 구간이 뒤쪽에 있지만, 100번째 데이터부터 Fault라고 라벨링을 하셨던 이유가 궁금합니다.

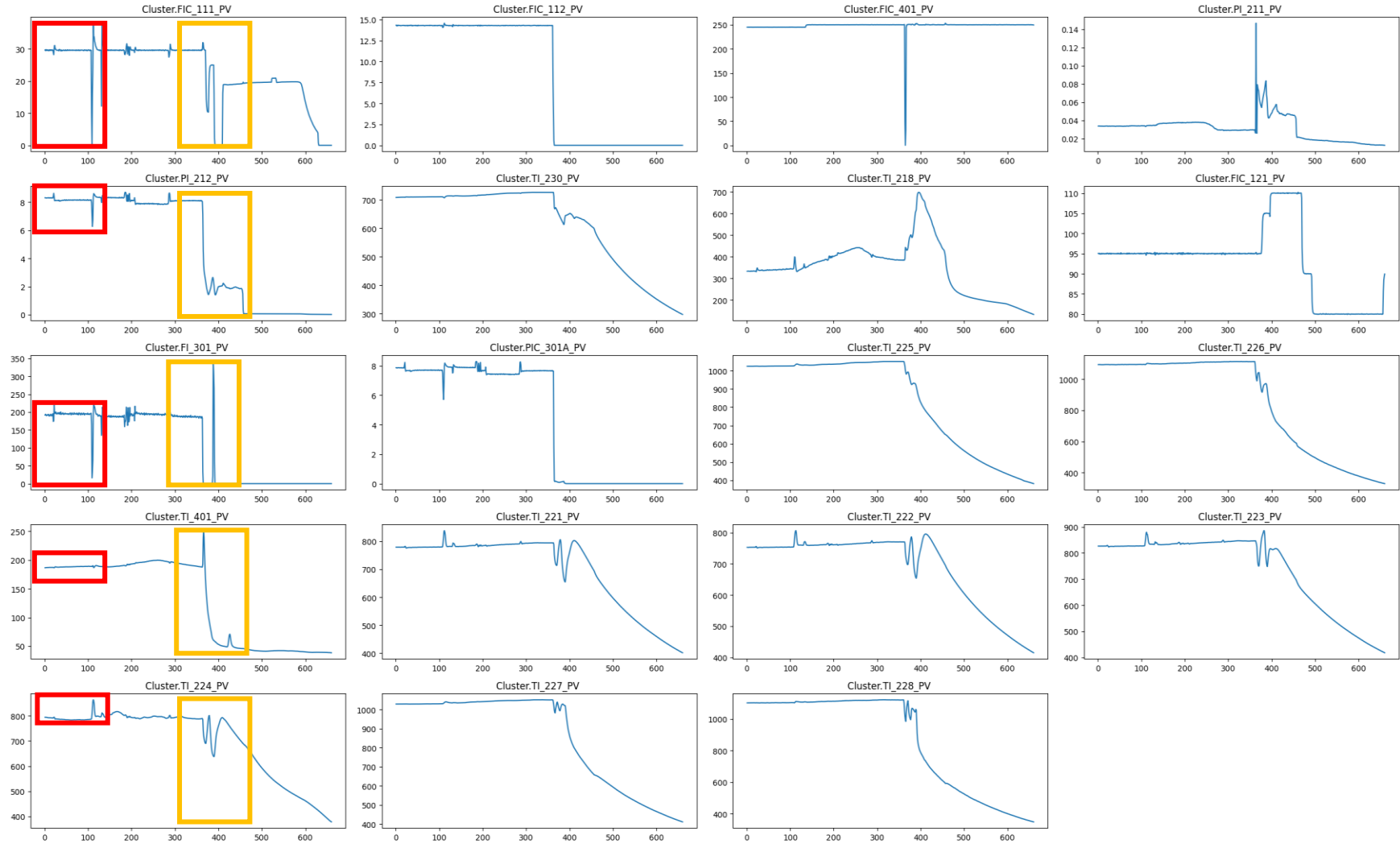


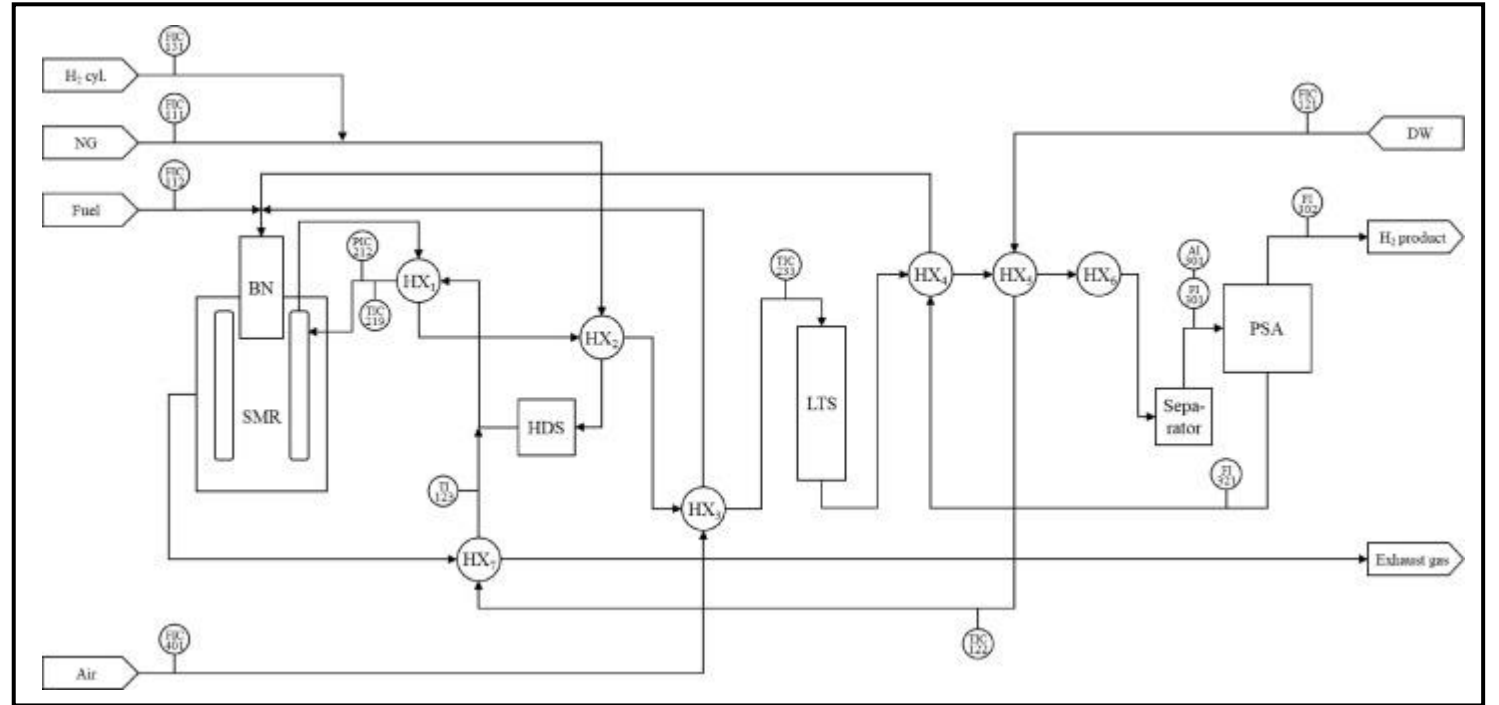
그림 1. 1차 데이터의 선택된 컬럼들의 데이터를 Plot한 결과

Requests

1. 공정 이상데이터를 담은 Raw데이터가 있다면 제공받고 싶습니다.
(zip파일안에 Raw데이터가 없었습니다..!)

2. 데이터프레임의 각 컬럼명이 이재원 박사님의 SMR onsite 시설에 관한 다른 논문의 Figure속에 있는 각 센서들의 이름들을 뜻하는 것으로 생각되는데요,
하지만 각 컬럼명이 공정 내에서 어떤 위치에서 구체적으로 무엇을 측정하는지는 알지못하는 상태입니다.

컬럼 명에 대한 구체적인 정보가 있다면 제공받고 싶습니다.



J. Lee, Machine learning-based energy optimization for on-site SMR hydrogen production

Answer

안녕하세요. 퇴근하고 확인해서 답변이 좀 늦었습니다ㅠ

1. 전반적으로 잘 이해하신 것 같습니다. 작성하신 내용에서 잘못된 부분은 없습니다. 다만, 이상데이터 라벨링 부분은 데이터 스케일링을 위해 정상 데이터 부분을 구분하기 위한 용도로 뒤의 모델링과는 상관이 없습니다. 이부분은 저번에 따로 설명을 안드렸던 것 같아 혹시나해서 말씀드립니다.

2. 작성하셨듯이 360번째부터 큰 변화가 있는 것이 맞고, 저도 연구 초반에 뒷부분 만을 Fault로 정의했었습니다. 그러나 빨간색 박스 부분이 정상데이터라고 하기엔 대부분의 변수에서 큰 변화(표준편차로 확인)가 일어나서 이재원교수님과 상의한 결과 빨간색 박스 부분에서 부터 문제(fault)가 일어나기 시작한 것으로 판단했습니다. 노란색 부분은 이미 반응기에 구멍이 난 것으로 fault라기 보단 사고(failure)로 정의했습니다.

3. Raw 데이터는 저도 어디갔는지 찾을 수가 없네요ㅠ 다시 한번 찾아보겠지만 확실하지 않아서 꼭 필요하시다면 이재원 교수님께 요청드려보면 좋을 것 같습니다.

4. 각 센서명과 매칭되는 공정도는 Data tag list.xlsx 파일에 그림으로 포함되어 있습니다. 공정에 관한 더욱 구체적인 이해가 필요하시다면 이재원교수님께 한번에 꼭 설명을 들어보시는 것도 좋을 것 같습니다.

박사님, 늦은시간인데도 답변해주셔서 감사드립니다!

1차와 2차 사고 모두 사고가 발생하기 조금 전 시간에 여러 변수에서 이상 현상이 발생하고 있으므로, 해당 이상현상이 사고와 연관성이 있을 가능성이 매우 높다고 판단할 수 있고, 따라서 이러한 이상현상을 탐지하는 것이 의미가 있는 것이군요.

현재 제가 raw 데이터를 볼 수는 없어서, 박사님께서 raw 데이터를 보셨을 때, 사고 직전이 아닌 평상시 상황에서도 여러 변수들이 한꺼번에 급격히 변동하는 현상이 종종 관찰되었는지 궁금합니다. 즉, 해당 현상이 정말 사고 직전에만 일어나는 특수한 현상인지, 아니면 일반적인 상황에서도 우연한 계기로 가끔씩 발생할 수도 있는 현상인지 알고 싶습니다.

감사합니다!

raw데이터라고 해서 전체 운전기간에 대한 건 아니고, 사고 전 2~3일치 정도를 더 포함해서 주셨던 걸로 기억합니다. 공정이 dynamic하게 움직이기 때문에 평상시 상황(steady-state)이라는 것을 따로 정의하기는 "어렵지만 여러 변수가 천천히 변화) 그렇다고 fault가 일어나면 무조건 failure로 발전하는한꺼번에 급격히 변동하는 현상"은 fault가 아닌 이상 일어나기는 힘듭니다. (정상적인 제어하에서는 것은 아니고 일반적으로 작은 fault들은 공정에서 스스로 제어됩니다.