

인공지능 모형을 이용한  
해양 유정 데이터셋 이상징후 감지

조명: 유정수루호

조원: 조수익, 이루안, 김경호, 정순주

## 2023년 지질자원 데이터 활용 및 인공지능 경진대회 최종 제안서

### 인공지능 모형을 이용한 해양 유정 데이터셋 이상징후 감지

조수익<sup>1)</sup>, 이루안<sup>2)</sup>, 김정호<sup>3)</sup>, 정순주<sup>4)</sup>

#### I. 서론

본 프로젝트는 인공지능 모형을 활용하여 해양 유정 데이터셋에서의 징후 탐지를 목적으로 한다. 이를 위해, 이상 사건 분류와 산업현장 도입을 위해 실시간 정상/이상 판별 모형을 구현하였다. 모형의 구현 및 평가를 위하여 해양 유정의 이상 탐지를 위해 공개된 3W데이터를 활용하였다. 프로젝트에서는 구체적으로 다음과 같은 3가지의 인공지능 모형을 제안한다. 이는 Masked late fusion CNN과 Stacked ML with TabNet, 마지막으로 Real time TabNet이다.

위 세 가지 모형을 통해 정확도가 높고, 실용적인 이상 징후 탐지 분류 모형 및 정확한 사후 이상 사건 분류를 목표로 한다.

#### II. 본론

##### 1. 데이터 분석 개요

###### (1) Dataset 정의

- 3W Dataset : 3개의 Offshore Oil-well에서 발생한 정상 및 이상 사건으로 이루어진 데이터셋

###### (2) Dataset 개괄

- 본 데이터셋은 A realistic and public dataset with rare undesirable real events in oil wells에 소개된 MAE project에 포함된 데이터셋
- Monitoramento de Alarmes Especialistas (MAE) project란, offshore naturally flowing wells에서 8가지 특정한 타입의 이상 상황(fault)을 짧은 시간 안에 더 나은 퍼포먼스로 분류해내는 자동화된 시스템을 개발을 목표로 하는 프로젝트로, Abnormal Event Management(AEM)의 영역에 포함
- AEM은 Responding to abnormal event로 아래 사항을 포함
  - ① Timely detection of an abnormal event

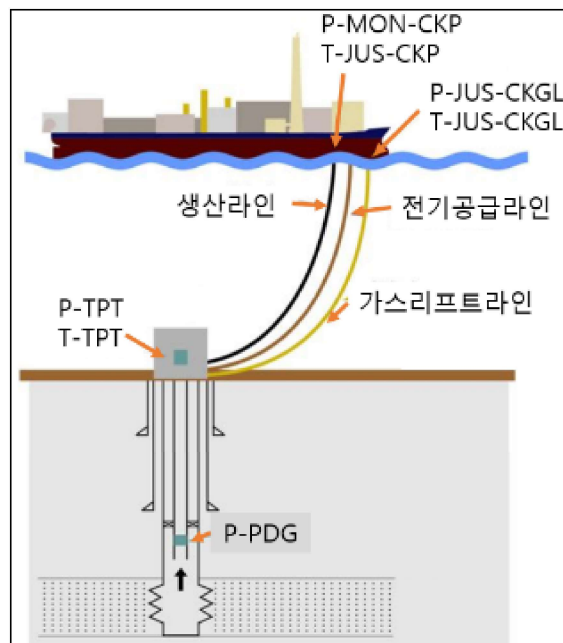
1) 모두의연구소/아이펠 (sik920405@gmail.com)

2) 모두의연구소/아이펠 (feelinppl@gmail.com)

3) 모두의연구소/아이펠 (khk172216@gmail.com)

4) 모두의연구소/아이펠 (9180jsj@naver.com)

- ② Diagnosing its root causes
- ③ Taking appropriate control decisions and actions to bring the process back to normal
- (3) Data 측정 장치 소개
  - 센서류
    - ① PDG(Permanent Downhole Gauge): 다운홀에서 측정되는 게이지 센서. 온도와 압력을 측정함.
    - ② TPT(Temperature and Pressure Transducer): Wellbore(산출구멍)에 위치한 변화기 센서. 온도와 압력을 측정함
  - 밸브류
    - ① DHSV(Downhole Safety Valve): 다운홀 안전 밸브
    - ② PCK(Production Choke): 생산 초크 밸브
- (4) Dataset Column 소개
  - P-PDG: 다운홀 게이지에서의 P(압력)
  - P-TPT: 압력 변화기에서의 P(압력)
  - T-TPT: 압력 변화기에서의 T(온도)
  - P-MON-CKP: 생산라인 상부의 초크밸브에서 관측된 P(압력)
  - T-JUS-CKP: 생산라인 상부의 초크밸브에서 관측된 T(온도)
  - P-JUS-CKGL: 가스리프트라인 상부의 초크밸브에서 관측된 P(압력)
  - T-JUS-CKGL: 가스리프트라인 상부의 초크밸브에서 관측된 T(온도)



[그림 1] 유정 데이터 관측 구조도

- (5) Dataset Event Type (Class) 소개
  - Class별 이상 발생 원인
    - ① Abrupt Increase of BSW: 갑작스러운 BSW(기초 침전물과 수분) 증가
    - ② Spurious Closure of DHSV: 문제상황이 아닐 때에 발생하는 비정상적 DHSV(지하안전밸브) 폐쇄
    - ③ Severe Slugging: 천연가스와 석유가 혼합되어 생산라인을 통과할 때 발생하는 현상. 유체의 흐름을 불안정하게 만들어 압력 변동, 유량 변동, 파이프 진동 등을 초래

- ④ Flow Instability: 불안정한 유동상태
- ⑤ Rapid Productivity Loss: 생산성 로스가 너무 급격히 떨어져서 극복할 수 없을 정도로 에너지가 낮아지는 경우
- ⑥ Quick Restriction in PCK: 생산 초크의 비정상적이고 급격한 제어
- ⑦ Scaling in PCK: PCK 장치에 침전물이 쌓인 상태
- ⑧ Hydrate in Production Line: Hydrate는 저온의 바다와 고압의 상태에서 가스 분자가 물 분자와 결합하여 가스가 얼음과 같이 결정화된 상태

[표 1] [1]Class 별 이상 상태를 확정하기 위한 관측 시간 (= window size)

Event Type	Window Size
Abrupt Increase of BSW	12H
Spurious Closure of DHSV	5min - 20min
Severe Slugging	5h
Flow Instability	15min
Rapid Productivity Loss	12H
Quick Restriction in PCK	15min
Scaling in PCK	72h
Hydrate in Production Line	30min-5H

(6) 3W Dataset의 구성

- Event > Instance > Variable > Observation 순으로 구성
- Sampling rate: fixed sampling rate 1Hz
- Instances: Well + Simulated + Drawn
  - ① Well: 실제 데이터 (실제 현장에서 추출된 실측 데이터, raw data 형태로 제공)
  - ② Simulated: 시뮬레이션 데이터 (시뮬레이터 OLGA에서 획득)
  - ③ Hand-drawn: 전문가들에 의해 작성된 데이터
- Label (이상 상태 구분)
  - ① Instance level: 이상 event type 판별
  - ② Observation level: 정상/비정상 (normal, faulty transient, and faulty steady state, null)
- Missing, Frozen, Unlabeled Data
  - ① Missing variables 4,947 (31.17% of all 15,872 variables)
  - ② Frozen variables 1,535 (9.67% of all 15,872 variables)
  - ③ Unlabeled observations 5,130 (0.01% of 50,913,215 observations of all 15,872 variables of all 1,984 instances)

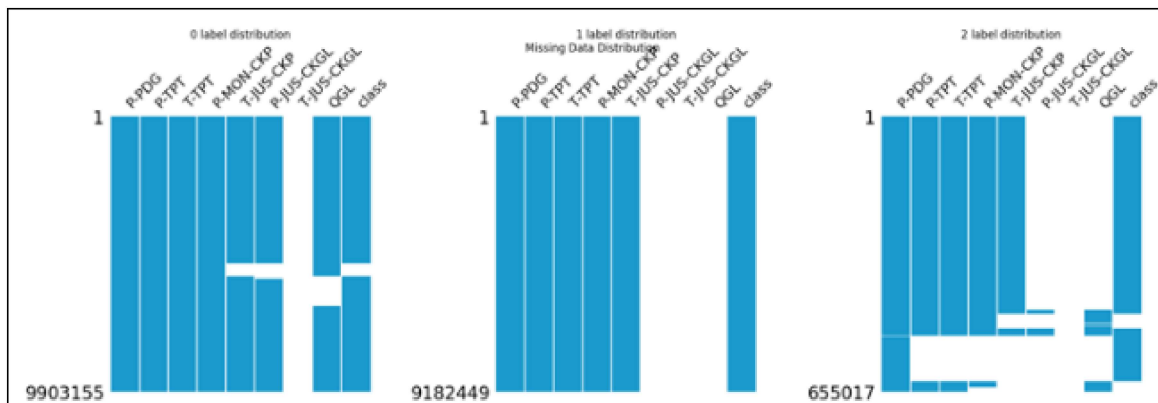
[표 2] [1]Class별 Instance 개수

Event Type	Window Size	Real	Simulated	Hand-drawn
Normal	-	597		
Abrupt Increase of BSW	12H	5	114	10
Spurious Closure of DHSV	5min - 20min	22	16	
Severe Slugging	5h	32	74	

Event Type	Window Size	Real	Simulated	Hand-drawn
Flow Instability	15min	344		
Rapid Productivity Loss	12H	12	439	
Quick Restriction in PCK	15min	6	215	
Scaling in PCK	72h	4		10
Hydrate in Production Line	30min-5H	3		

## (7) 데이터 분석 결과

- 결측치를 고려하여 P-JUS-CKGL, T-JUS-CKGL, QGL 칼럼 제외



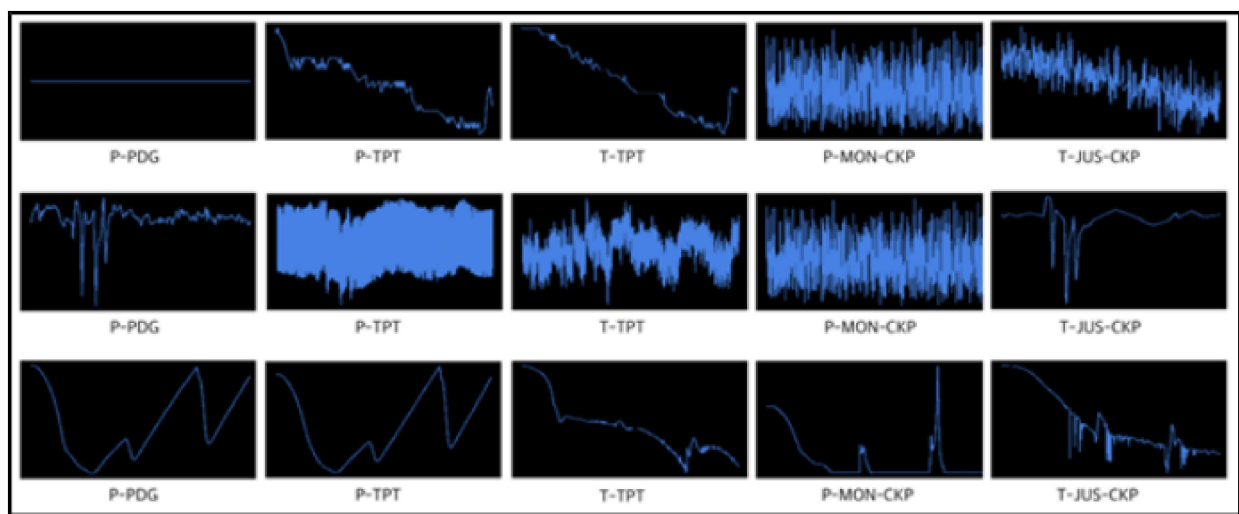
[그림 2] 클래스/칼럼별 결측치 확인 그래프 中 (0-2 event)

## 2. 인공지능 모형

### (1) Masked late fusion CNN

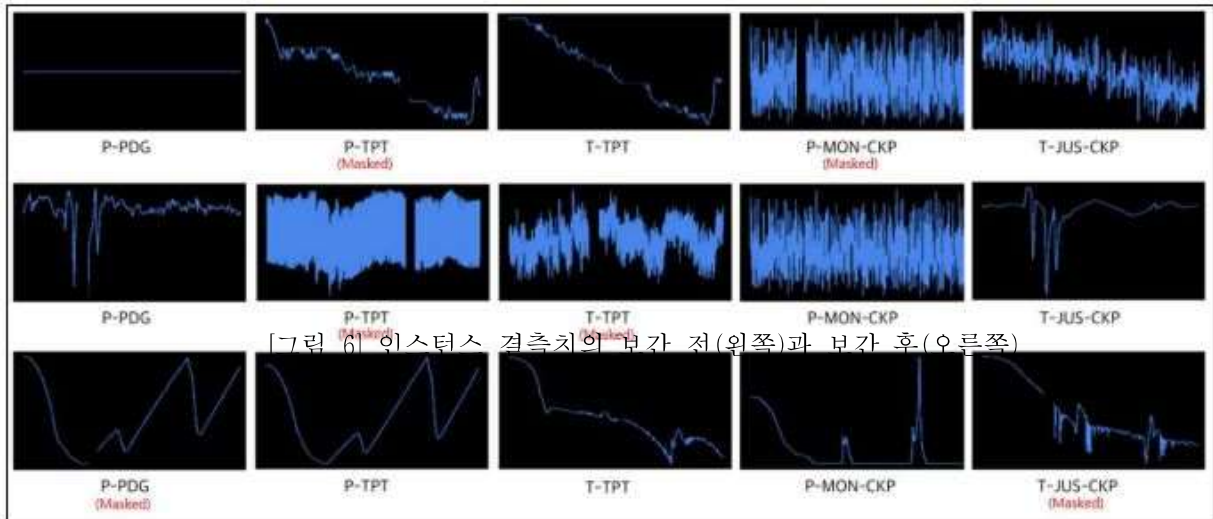
- 데이터 구성

#### ① Plot Image



[그림 3] Plot Image 예시

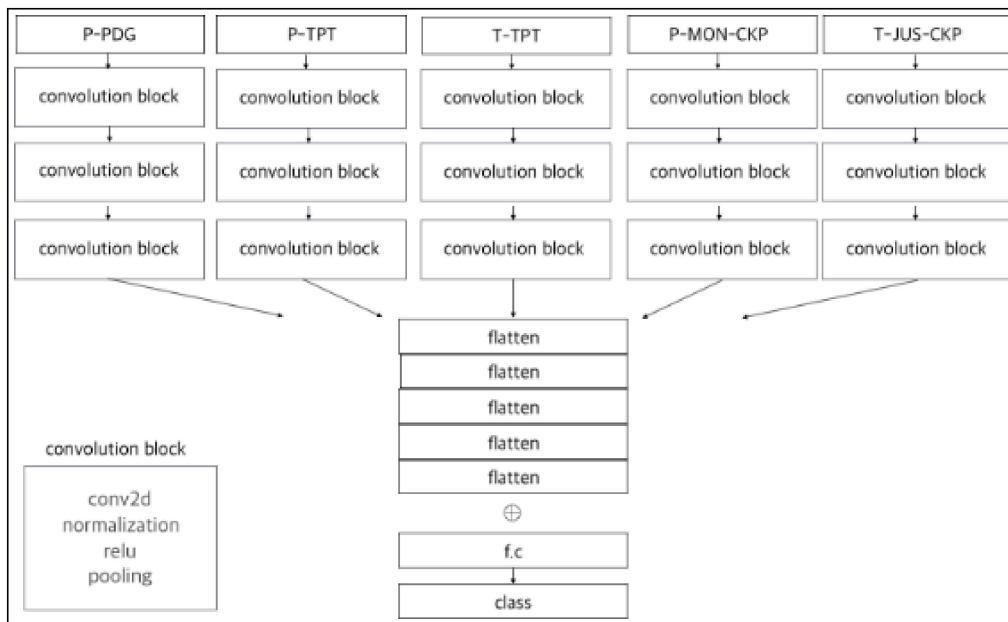
- ② Masking : 1, 2, 5, 6, 7 Classes의 경우, Well-data만 사용하여 칼럼 2개를 임의 추출 후 3% 마스킹 적용



[그림 4] Masked Plot Image 예시

## - 모형 구조도

### ① Multi-Input CNN



[그림 5] Late fusion CNN model architecture

### (2) Stacked ML with TabNet

#### - 데이터 구성

##### ① 결측치 보간: Linear 방식으로 10% 미만의 결측을 보간

##### ② 차분 및 PCA 데이터 추가: Feature 다양화를 통해 클래스 간 Dependency 해소

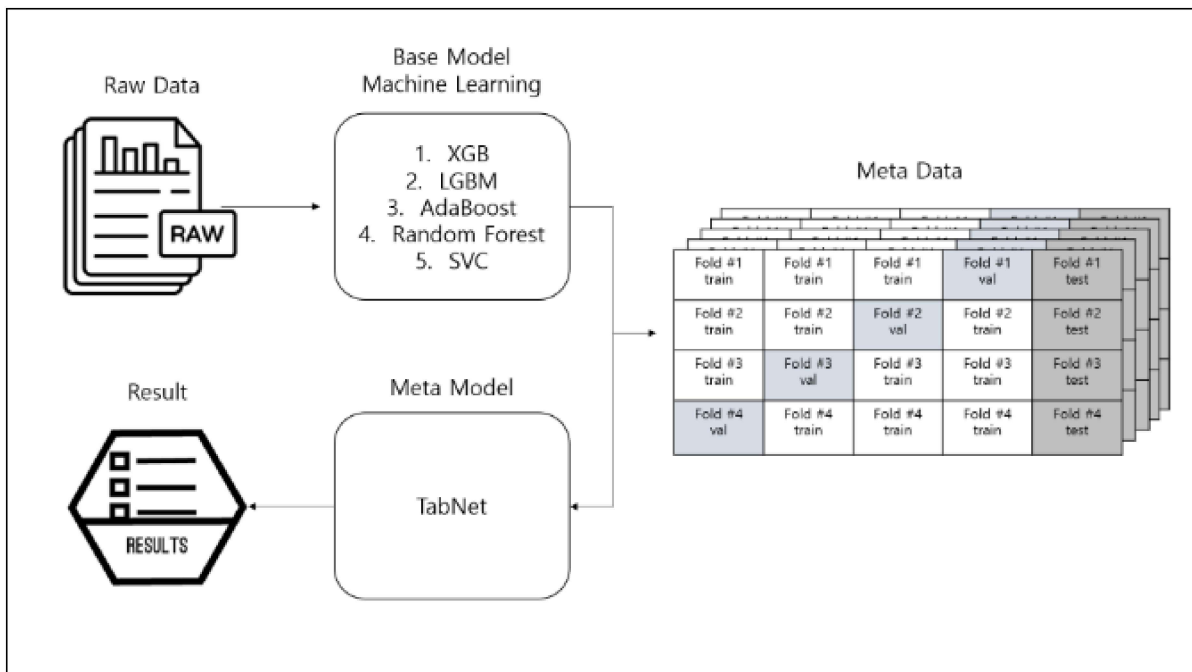
	P-PDG	P-TPT	T-TPT	P-MON-CXP	T-JUS-CXP	id_label	class	observation_nums		P-PDG	P-TPT	T-TPT	P-MON-CXP	T-JUS-CXP	id_label	class	observation_nums
0	489	489	489	512	490	0	0	8431490	0	0	0	0	0	0	0	0	8431490
1	0	0	0	0	0	922	0	9158432	1	0	0	0	0	0	0	922	9158432
2	0	0	0	0	0	140	0	486452	2	0	0	0	0	0	0	140	486452
3	581	581	581	674	544	0	0	4034079	3	0	0	0	0	0	0	0	4834079
4	758	1067	1056	672	1126	0	0	2462076	4	0	0	0	0	0	0	0	2462076
5	1128	1128	1124	1133	1132	12500	0	13161262	5	0	0	0	0	0	0	12500	13161262
6	0	0	0	0	0	1056	0	56578	6	0	0	0	0	0	0	1056	56578
7	362	362	362	362	362	5167	0	2885548	7	0	0	0	0	0	0	5167	2885548
8	0	0	0	0	0	0	0	2186920	8	0	0	0	0	0	0	0	2186920

[그림 8] Real time TabNet 입력데이터 예시

③ 시계열 대표화: 다섯 등분 후 단위별 평균, 표준편차 계산

- 모형 구조도

- ① ML Stacking으로 메타 데이터 생성: K-Fold 방식 사용
- ② TabNet을 통한 최종 결과 예측

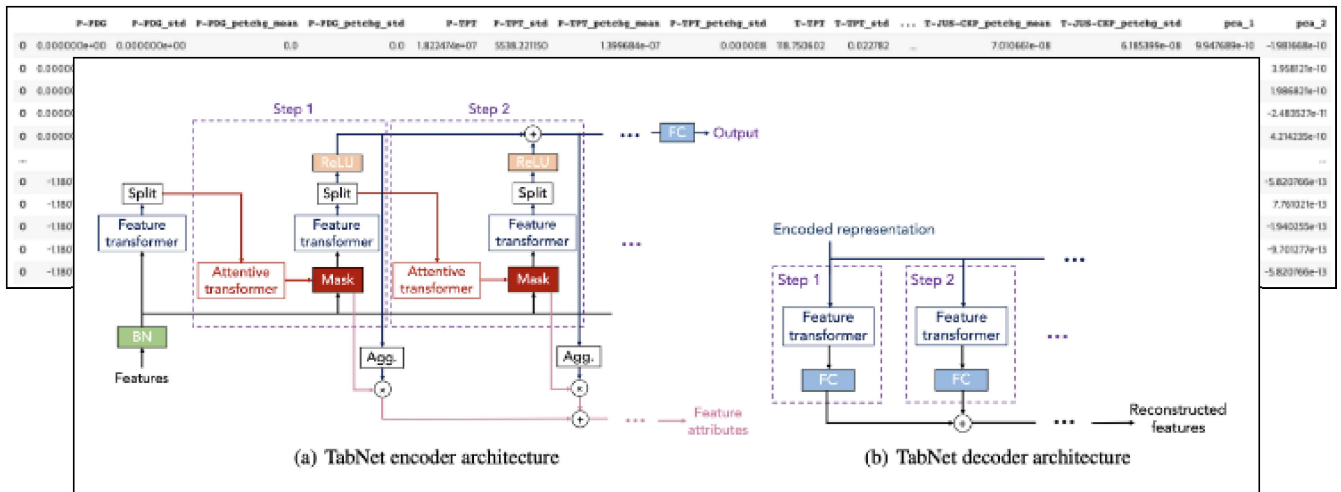


[그림 7] Stacked ML with TabNet 도식

(3) Real time TabNet

- 데이터 구성

- ① 차분 및 PCA 데이터 추가: Feature 다양화
- ② Window 적용: 실시간성 확보를 위한 짧고 일정한 Window 적용



[그림 9] [2]TabNet 구조도

- 모형 구조도

### III. 결론

#### 1. 결과 및 성능 평가

##### (1) Masked late fusion CNN

- ① Accuracy: 99.0%
- ② 평가: 높은 Accuracy의 모형으로, 보다 정확한 이상 사건 분류
- ③ [Code link](#)

##### (2) Stacked ML with TabNet

- ① Accuracy: 94.6%
- ② 평가: DL을 사용한 CNN 기반 모형보다 경량화된 ML 기반의 모형을 사용과 동시에 준수한 성능 구현 및 시간 단축
- ③ [Code link](#)

##### (3) Real time TabNet

- ① Accuracy: 96.2%
- ② 평가: 주어진 데이터셋에 국한되지 않은 실제 산업현장에서의 실시간 데이터를 활용하여 이상 사건 분류 가능
- ③ [Code link](#)

#### 2. 개발 모형의 활용방안

- (1) 실시간 이상 징후 탐지: 산업현장에서 실시간으로 수집되는 데이터를 Real time TabNet 모형에 적용, 이상징후 여부를 우선 탐지
- (2) 사후 이상 클래스 분류: 이상징후 탐지 후, 정확도 높은 Masked late fusion CNN 모형 혹은 Stacked ML with TabNet 모형을 활용하여 이상 클래스를 분류



#### 참고문헌

- [1]. Ricardo E, Celso J.M. A realistic and public dataset with rare undesirable real events in oil wells. Journal of Petroleum Science and Engineering. Volume 18., October 2019, 106223
- [2]. <https://arxiv.org/pdf/1908.07442.pdf>