

# HiMSEN 엔진에서 측정한 안다미로 아롬에 대한 분석

Team - 3Miroa



# 목차



Intro



데이터 전처리



탐색적 데이터 분석



알고리즘 적용



기능 및 성능 평가



Discussion



References



Q&A



# Intro

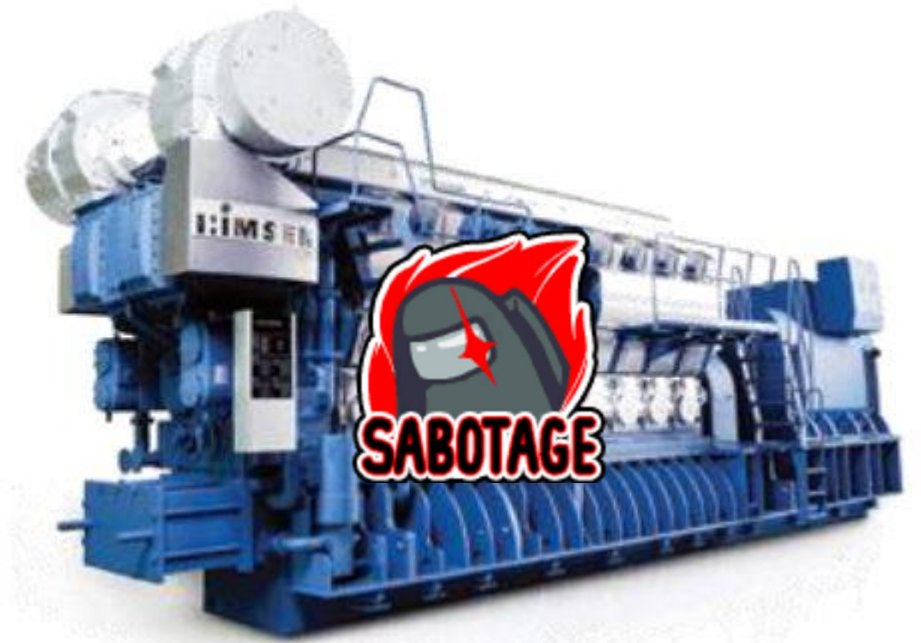


우주선을 고장내는 크루원(임포스터)를  
잡고 임무를 수행하는 크루원들

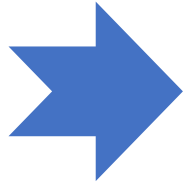
임포스터가 우주선 내부 시설을  
고장내는 행위인 사보타지



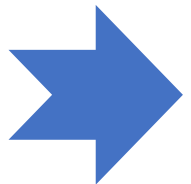
빅데이터를 통해 힘센엔진에 대해  
파악하고 고장으로부터 지켜낸다!



# 방향성



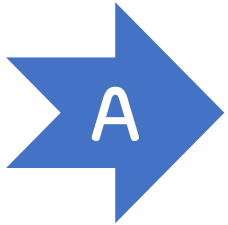
각 시스템별로 이상을 진단하는 기법을 제안  
➤ 문제가 생긴 시스템에 선별적, 선제적대응 가능



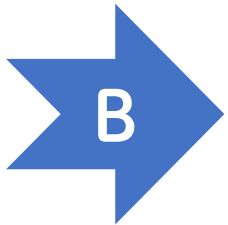
네 가지 시스템(엔진, T/C, Air Cooler, Cooling Water)  
에 대해 개별적 건강지수와 통합 건강지수를 제안  
➤ 엔진의 종합적 상태를 진단하는 알고리즘 제안

# 문제 정의

$$\text{복합 문제} = (\text{비지도학습} + \text{지도학습}) \times (\text{분류} + \text{회귀})$$



라벨링 되지 않은 개별 시스템의 고장 여부 판단,  
각각을 점수화

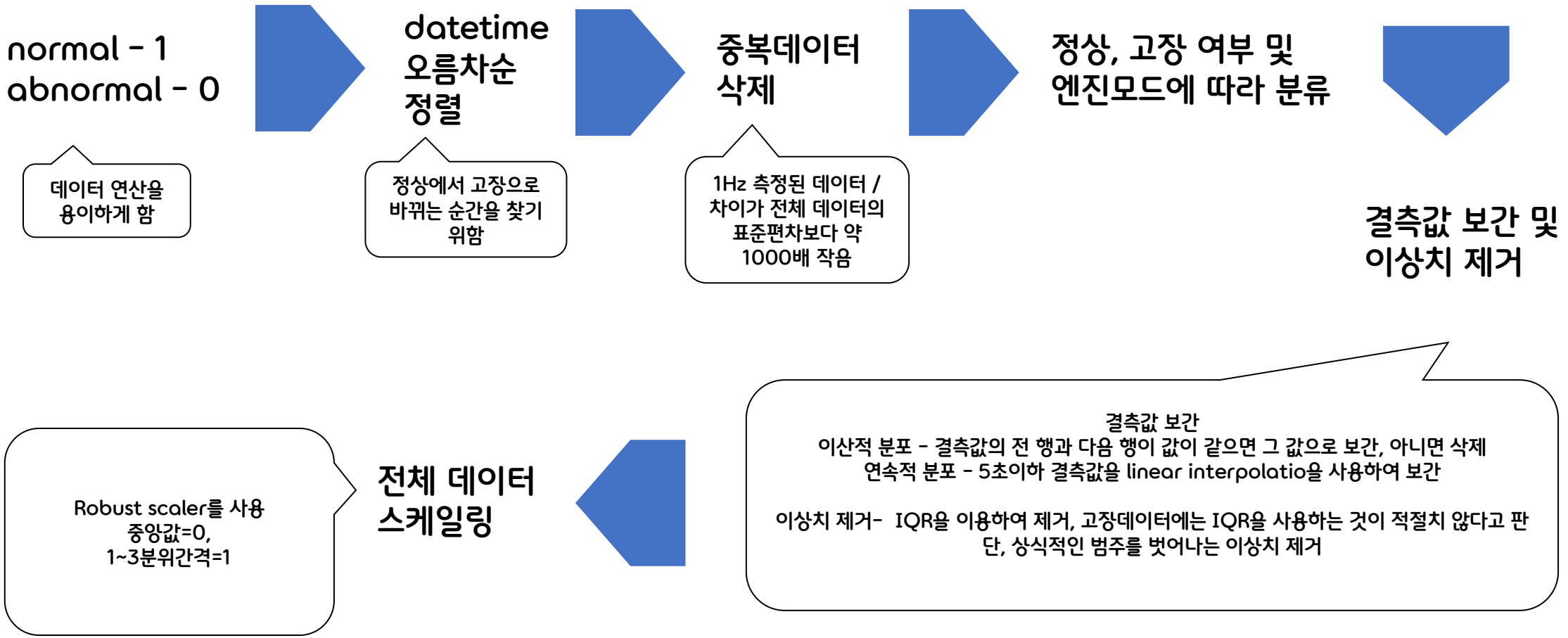


A에서 판단한 개별시스템의 결과를 종합하여 낸 결과와  
실제 전체 시스템의 고장 여부와 일치 여부 확인 및 점  
수화



# 데이터 전처리

## 순서도



사용자

## 전체 시스템 개념도

fuel mode, engine speed

연료펌프

Base duration  
gas admission  
Gas temp

가스

Gas press,  
Engine inlet

연료펌프

FO Pressure  
Engine Inlet

FO Temperature  
Engine Inlet

디젤

동력

크랭크

Engine Load  
percent

전력

Main bearing  
temp average

Combustion

Engine speed

IMEP

Pmax

DVT control air press

냉각수

Charge Air press,  
Engine Inlet

Charge Air temp,  
Engine Inlet

공기

Air temp, TC inlet

Exh. Gas temp  
average

배기가스

Exh. Gas temp,  
TC Inlet A Bank

TC

TC speed

Waste Gate Valve  
position command

배기가스

Exh. Gas temp,  
TC Outlet A  
Bank

냉각수

HT CW temp, Engine Inlet

냉각수

Air Cooler

LT CW press,  
Air Cooler Inlet  
LT CW temp,  
Air Cooler Inlet

냉각수

LT CW  
temp, Air  
Cooler  
Outlet

Cooling  
water

HT CW temp, Engine Outlet

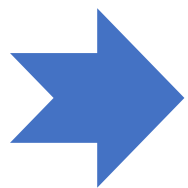
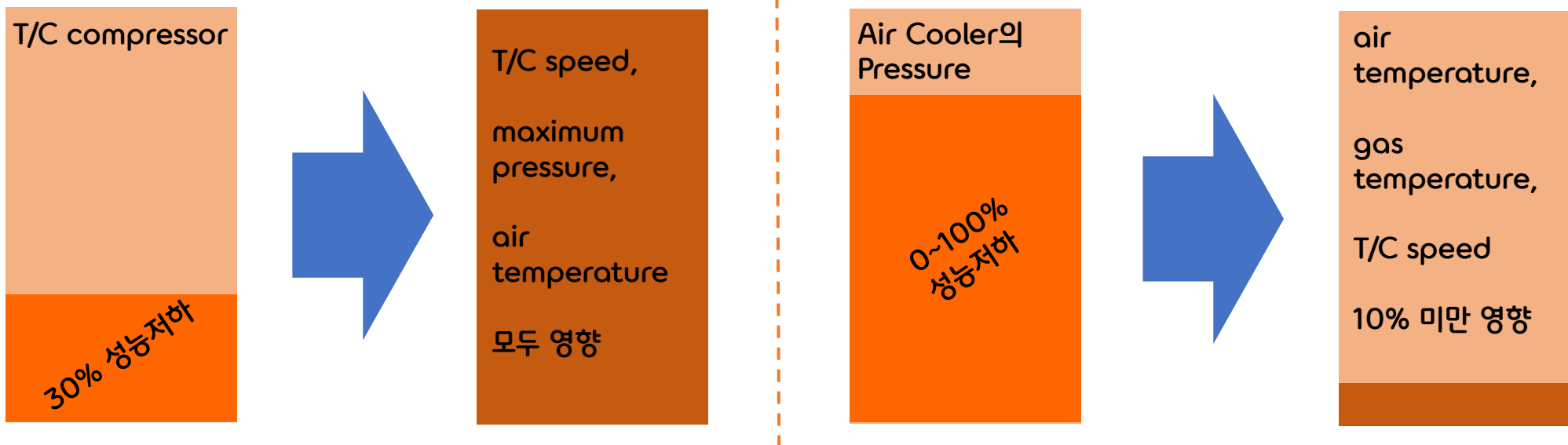




# 탐색적 데이터 분석

전반적인 고장의 유형과 변수간 상호작용 방식

= D.T.Hountalas의 연구에서 Air Cooler, Combustion, Turbo Charger의 고장 유형과 상호작용 방식을 참고하여 결정



T/C의 고장여부를 T/C speed, maximum pressure, air temperature의 변동을 통해 역추적 할 수 있는 요소로 생각할 수 있고, air cooler의 pressure는 엔진의 다른 부분 및 고장 여부에 크게 영향을 주지 않는 독립적인 요인으로 간주

## 데이터 분류 및 새로운 변수 생성

데이터 종류	ㄱ	ㄴ	ㄷ	ㄹ
특징	이산적	특정 범위 내 분포	상호 연관성, 종속적	연관성을 바탕으로 유도
예시	Base duration gas admission	베어링 온도, 냉각수 온도 등	엔진로드, TC speed, exh gas TC outlet 등	엔진로드/IMEP

ㄱ) 고정된 값 가짐. 예를 들어 연료모드 2의 경우, Base duration gas admission 센서 값은 항상 0.

➡ 학습 의미 X. 특정 값이 아닌 경우 고장으로 분류.

ㄴ) 특정 범위를 벗어나지 않으면 안전. 표준정규분포를 사용해 정상 데이터 범위 밖에 해당하는 값은 고장으로 판단.

ㄷ) 대부분 TC와 Combustion에 관련된 센서로 부분적인 연관성을 갖는 데이터가 포함. 뚜렷한 선형적 관계나 상관관계 X, 복잡한 상관관계 가짐.

➡ 신경망 알고리즘을 통해 학습.

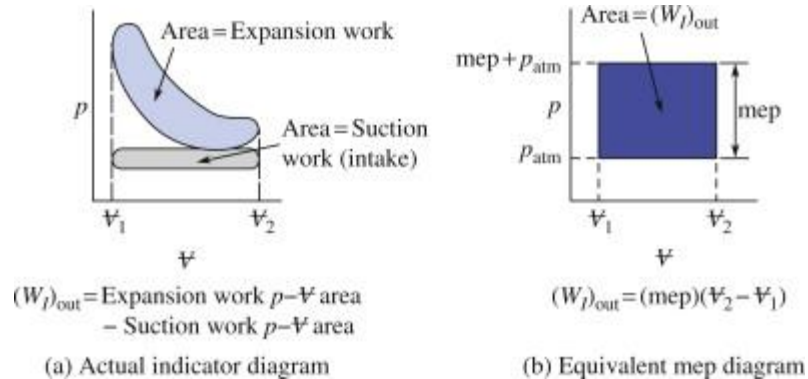
ㄹ) 내연기관의 특성을 사용하여 유도된 값.

## 변수분류표

Mode2	Mode3	시스템	변수명
ㄱ	ㄴ	Combustion	Base duration gas admission
		T/C	Waste Gate Valve position command (CV26)
ㄴ	ㄴ	CoolingWater	Main bearing temp average
		AirCooler	LT CW press, Air Cooler Inlet
			LT CW temp, Air Cooler Inlet (TE71)
			LT CW temp, Air Cooler Outlet (TE72)
		CoolingWater	HT CW temp, Engine Inlet (TE75)
			HT CW temp, Engine Outlet (TE76)
			Main bearing temp average
		Combustion/AirCooler	Charge Air press, Engine Inlet (PT21X)
			Charge Air temp, Engine Inlet (TE21X)
ㄷ	ㄷ	T/C	Engine Load percent
			Air temp, TC inlet (TE29)
			Exh. Gas temp, TC Outlet A Bank (TE27-A)
			TC speed A bank (SE14-A)
		Combustion	Engine Speed 1 (SE11)
			Exh. Gas temp average
			IMEP average
			Pmax average
		T/C, Combustion	DVT control air press (PT43)
			Exh. Gas temp, TC Inlet A Bank (TE26-A)
	ㄴ	Combustion	FO Pressure Engine Inlet
			FO Temperature Engine Inlet
ㄴ	ㄷ	Combustion	Gas press, Engine Inlet (PT87)
			Gas temp (TE81)
ㄹ	ㄹ	Combustion	(Load-b)/(IMEP+1)
			(Load)/((IMEP+1)x(Engine speed +1))
			(Pmax average-b)/(IMEP+1)

## ≡ 데이터 유도과정

### 1 Boundary work과 평균 유효 압력의 관계



$$W = mep(V_2 - V_1)$$

$$\dot{W} = mep(n)(V_2 - V_1) \left( \frac{N}{C} \right)$$

### 2 V<sub>2</sub>-V<sub>1</sub>이 일정하지 않으면 결함이 있는 것으로 판단

a. 1초 단위로 Engine Load Percentage를 시간에 대해 적분함

$$\frac{\text{Engine Load percent}}{IMEP} = k \times (V_2 - V_1)$$

b. 해당되는 센서값들을 대입

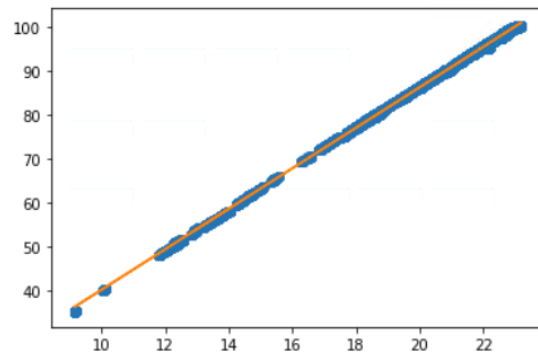
$$\frac{\text{Engine Load percent}}{IMEP \times \text{Engine Speed}} = k_2 \times (V_2 - V_1)$$

## ㄹ 데이터 유도과정

- 3 a. 분모(IMEP)가 0이 되는 순간을 방지하기 위해 IMEP에 대기압인 1을 더해 gauge pressure로 변환.

선형회귀분석을 진행한 뒤 선형회귀식의 기울기를 새로운 변수로 사용

- b. 분모(Engine Speed)가 0이 되는 것을 막기 위해 1 더함



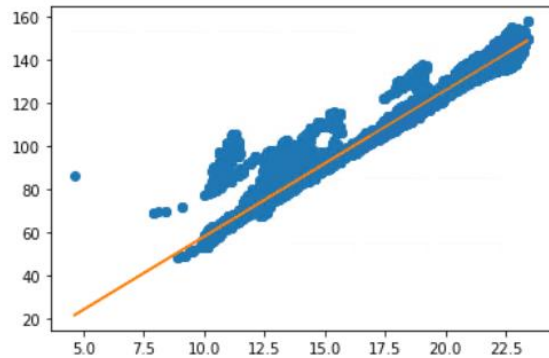
- 4 C는 상수로 4행정기관에선 2이고, n은 기관의 분당 회전수(RPM), n은 실린더의 수를 의미한다.

위의 수식을 정리하고 해당되는 센서값들을 대입

$$\frac{\text{Engine Load percent}}{\text{IMEP} \times \text{Engine Speed}} = k_2 \times (V2 - V1)$$

5

Pmax와 IMEP 비율을 새로운 변수로 정의



## 기존 연구와의 차이

### 1. 주파수 계열 방식 적용 어려움

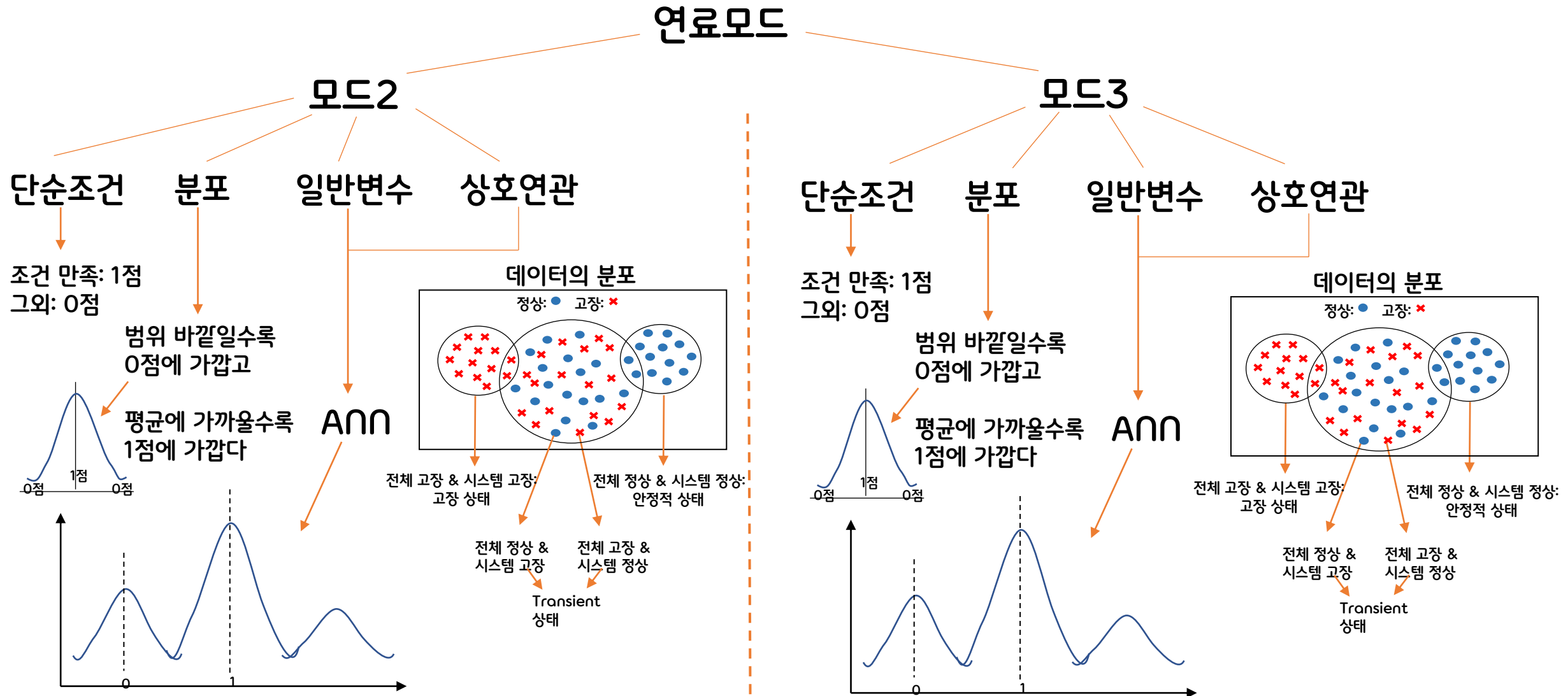
- Nyquist Theorem

### 2. Feature Selection Method(적절한 크기의 묶음으로 나누어 특징을 추출하는 방식) 적용의 어려움

- 낮은 sampling rate에서 발생하는 비정규성
- 적절한 p-value 선정의 어려움



# 일고리즘 적용





# 데이터 종류별 고장 분류

## ㄱ) 단순 조건

- 특정 엔진모드에서는 특정 값만을 가져야하므로 단순 조건식을 통해 특정 값에 해당되면 정상(1) 그렇지 않으면 고장(0)으로 분류

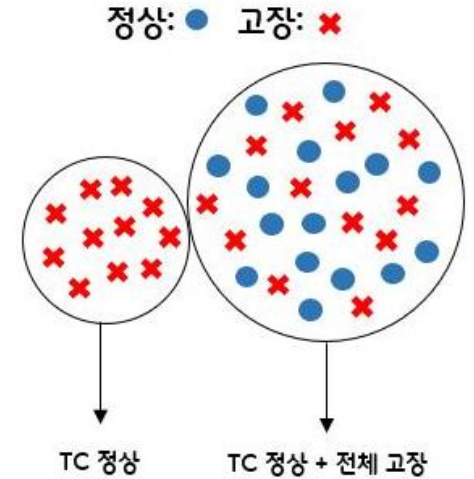
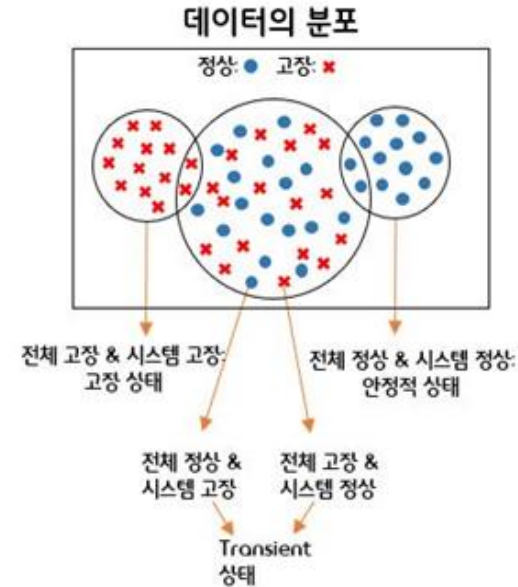
## ㄴ) 통계적 기법 사용

- 데이터 표준화 후, Z점수를 이용해  $Z > 0$ 일땐,  $Z' = Z / \max(Z)$ ,  $Z < 0$ 일땐,  $Z' = Z / \min(Z)$ 으로 하여 범위를 0~1로 설정한다. 그 후,  $Z'' = 1 - Z'$ 로 설정하여 최종점수는 평균에 가까울수록 1에 가깝고 평균에서 멀어질수록 0에 가깝게 설정.

## ㄷ, ㄹ) ANN 사용

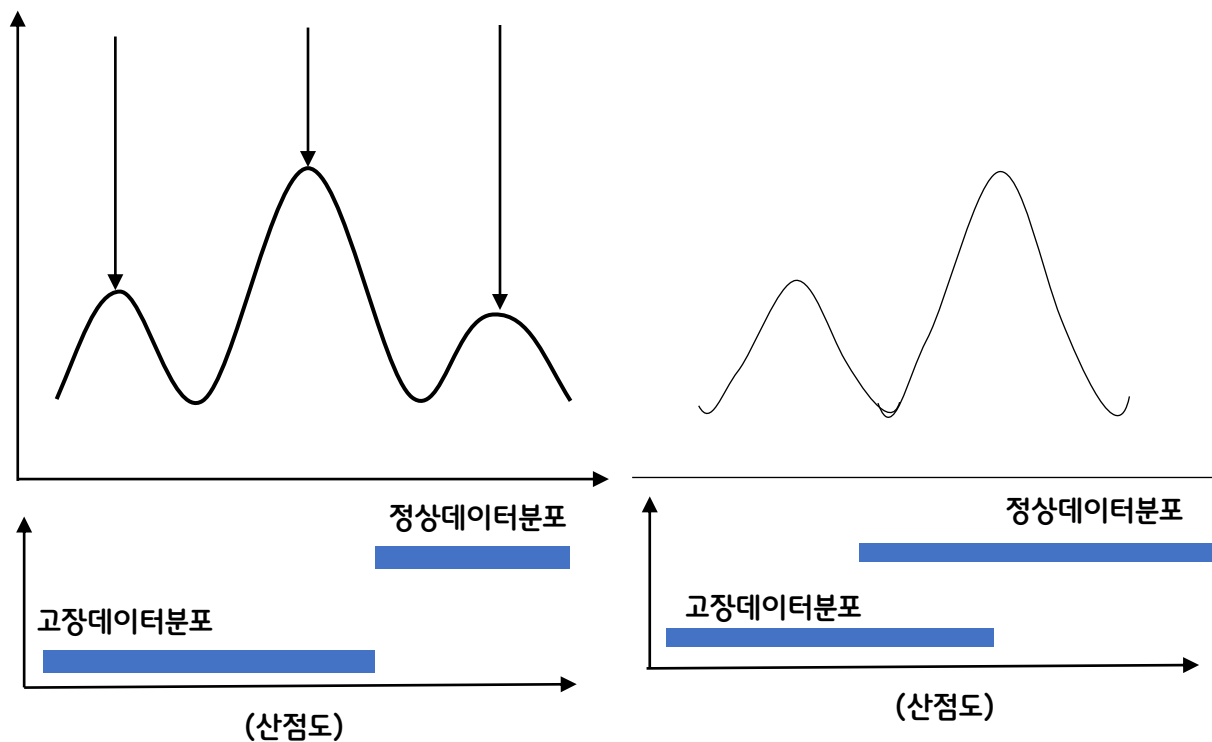
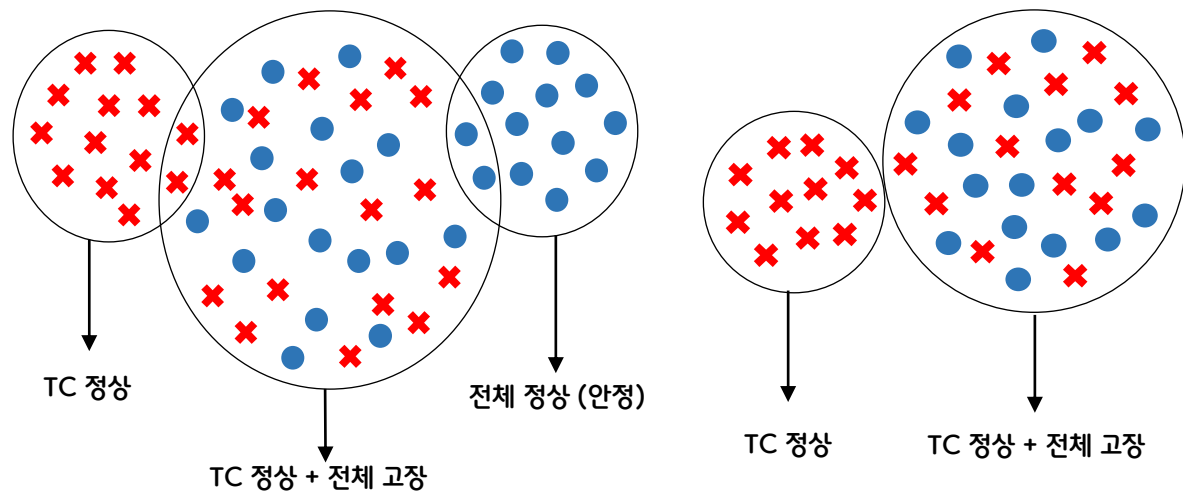
- 머신러닝 / 딥러닝 - 고장 분류 가능, 원인 찾기 불가.  
Clustering - 고장 데이터 분포의 다양성으로 인해 결과 신뢰성 보장 X. 점수화 어려움.
- 텐서플로 ANN 알고리즘 - 정상 1, 고장 0으로 하는 회귀방식 사용.  
Input - ㄷ, ㄹ 데이터의 feature / output - 전체 시스템의 정상(1), 고장(0).  
train:validation:test = 6:2:2로 하여 과적합 방지.

## 예상되는 데이터의 두 가지 분포



# ANN 사용 이유

정상: ● 고장: ✕

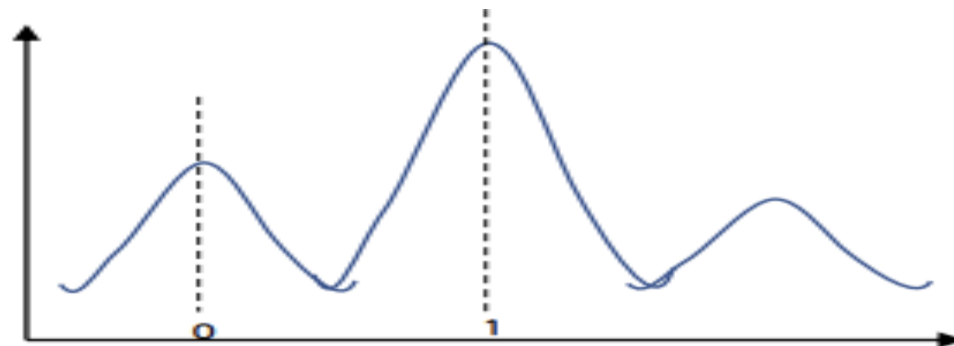


전체 시스템에선 고장이지만 특정 시스템에선 정상인 고장데이터의 영향 최소화 가능.

ANN의 회귀방식을 사용한다면 0~1사이의 값을 output으로 반환하게 되고 1에 가까울수록 정상, 0에 가까울수록 고장으로 판단.

일반적인 분류와는 달리 특정 시스템 내에서는 정상과 차이가 없는 고장데이터도 많은 비율을 차지하기에 ANN은 정상 값에 대해서는 대부분 1에 가까울 만큼 높은 수치가 아닌 0.6~ 0.5 정도의 낮은 값을 냄.(좌측 그림 참고)

시스템 별로 최종 건강지수 도출할 때에는 이 점을 보완하여 앙상블 과정에서 0~1로 나오도록 스케일을 조정.  
(하단 그림 참고)



# ANN 구성 및 앙상블을 통한 개별 시스템 알고리즘 보완

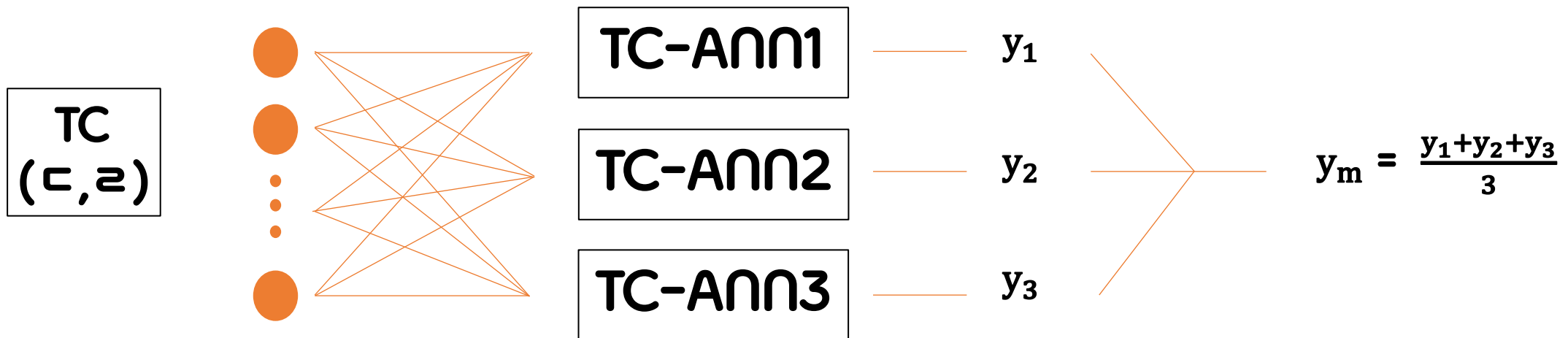
T/C와 Combution에 적용. / Epoch 수: 100 / Hidden layer 수: 4 / Sequential mode.

n개 차원 데이터에 대해 2n+1개 이상 노드 사용 (Kolmogorov rule): 각 시스템별 입력 데이터 차원에 맞게 노드 수 조절.  
mode 2와 mode 3에서 4개 차원의 input을 가진 T/C는 9개,  
mode 2에서 12개, mode3에서 9개 차원을 갖는 Combustion은 각각 25개와 19개 노드를 갖도록 조정.

알고리즘 핵심: 과적합 방지, (보편적인 경우라면 매우 낮은) 적절한 정확도에서 학습을 멈추도록 하여야 라벨이 고장이지만

→ 특정 시스템에서는 정상으로 동작하는 데이터를 오판하지 않음.  
validation과 10%이상 차이가 날 시 과적합방지를 위해 조기에 학습을 종료하도록 지정.

‘특정 시스템에서는 정상범주에 있는 고장데이터’를 특정시스템에서도 모두 고장으로 처리해버릴 가능성을 배제하기 위해 soft-voting 기법을 참고하여, 특정 시스템에 대해서만 학습한 다수의 ANN의 output들을 평균 낸 후 다시 0~1로 스케일링하여 특정 시스템의 최종 output으로 결정.(하단 그림 참고)





# 기능 및 성능 평가

## 성능 평가

양상블을 통해 추정을 종합하여 전체 시스템에 대해 고장진단을 내렸을 때의 결과가 일치하는지 검증.

전체 시스템 중 하나라도 고장이 생긴 케이스와 데이터의 abnormal 레이블이 높은 일치율을 보이고,  
모든 시스템에 고장이 없는 케이스와 데이터의 normal 레이블이 높은 일치율을 보임

➡ 전체시스템에 대한 고장 진단의 정확도가 높다고 판단

양상블은 개별 시스템의 output들의 minimum을 반올림해 비교

# 최종 모델: 종합 건강지수 도출

시스템별로 점수정리

CW: 단순조건, 분포 → 평균\*100 = CW 건강지수

AC: 단순조건, 분포 → 평균\*100 = AC 건강지수

TC: 단순조건, 분포 → 평균\*100 = TC 건강지수

Combution: 단순조건, 분포 → 평균\*100 = Combution 건강지수

종합 건강 지수 =

$$\frac{(CW \text{ 건강지수} \times CW \text{ 변수 수}) + (AC \text{ 건강지수} \times AC \text{ 변수 수}) + (TC \text{ 건강지수} \times TC \text{ 변수 수}) + (Combution \text{ 건강지수} \times Combution \text{ 변수 수})}{\text{총 변수 수}}$$

# 항목별 고장 진단 결과

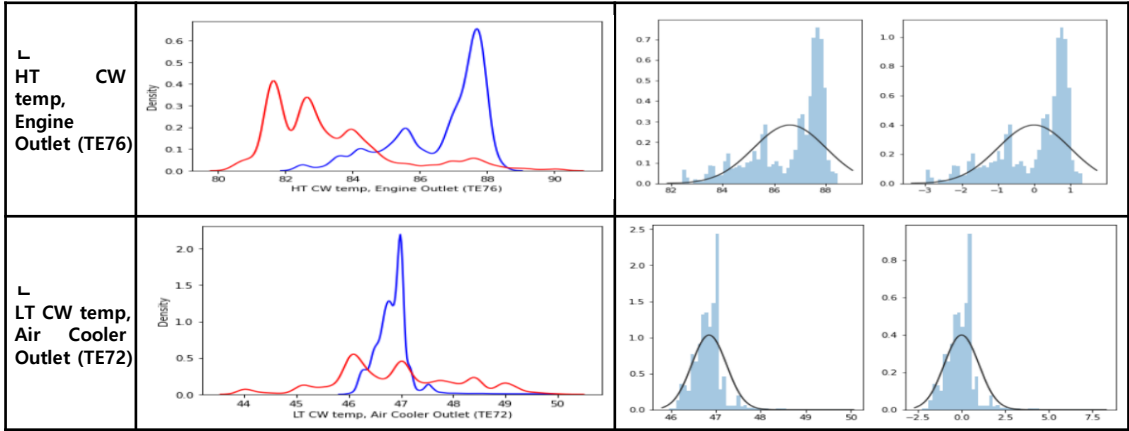
## ㄱ, ㄴ 데이터

	밀도그래프 (정상 - 파랑 / 비정상 - 빨강)	표준정규그래프 (좌 - 치환 전 / 우 - 치환 후)
ㄱ LT CW press , Air Cooler Inlet (PT71)		없음
ㄴ Base duration gas admission		없음
ㄴ Waste Gate Valve position command (CV26)		없음
ㄴ Main bearing temp average		
ㄴ LT CW temp , Air Cooler Inlet (TE71)		

ㄴ HT CW temp, Engine Inlet (TE75)		
ㄴ Charge Air press, Engine Inlet (PT21X)		
ㄴ Charge Air temp, Engine Inlet (TE21X)		
ㄴ Gas press, Engine Inlet (PT87)		
ㄴ Gas temp (TE81)		

# 항목별 고장 진단 결과

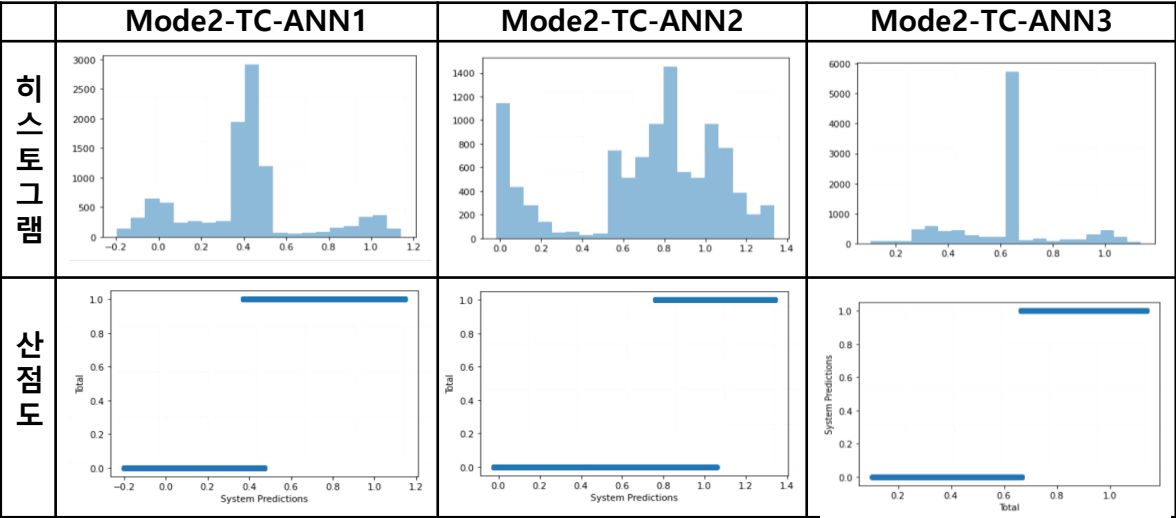
## ㄱ, ㄴ 데이터



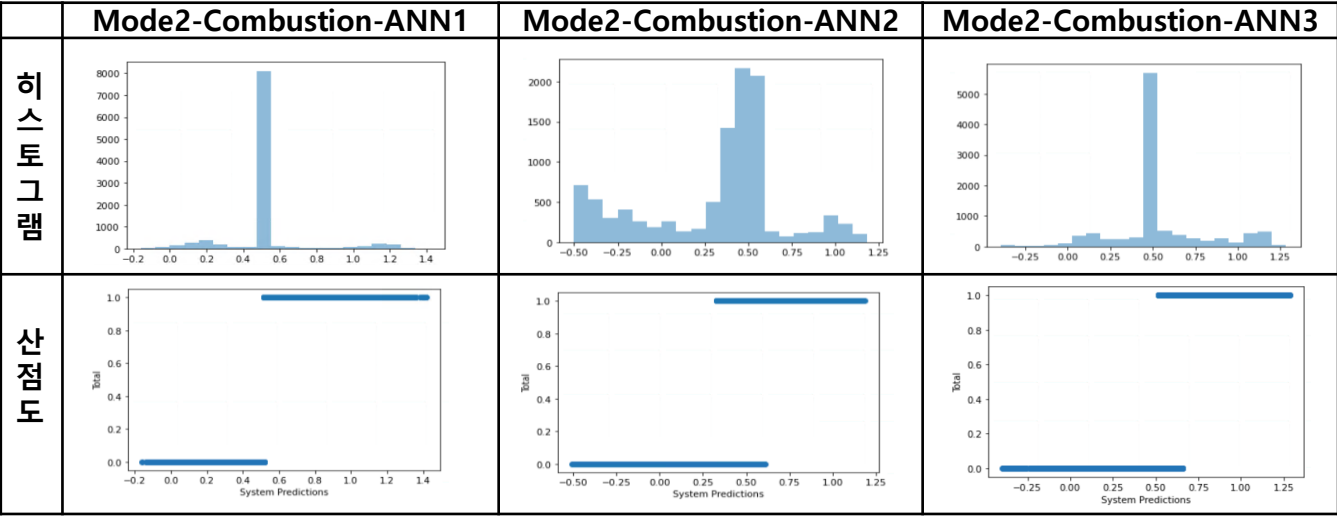
# 항목별 고장 진단 결과

## ㄷ, ㄹ 데이터

Train-mode2-TC



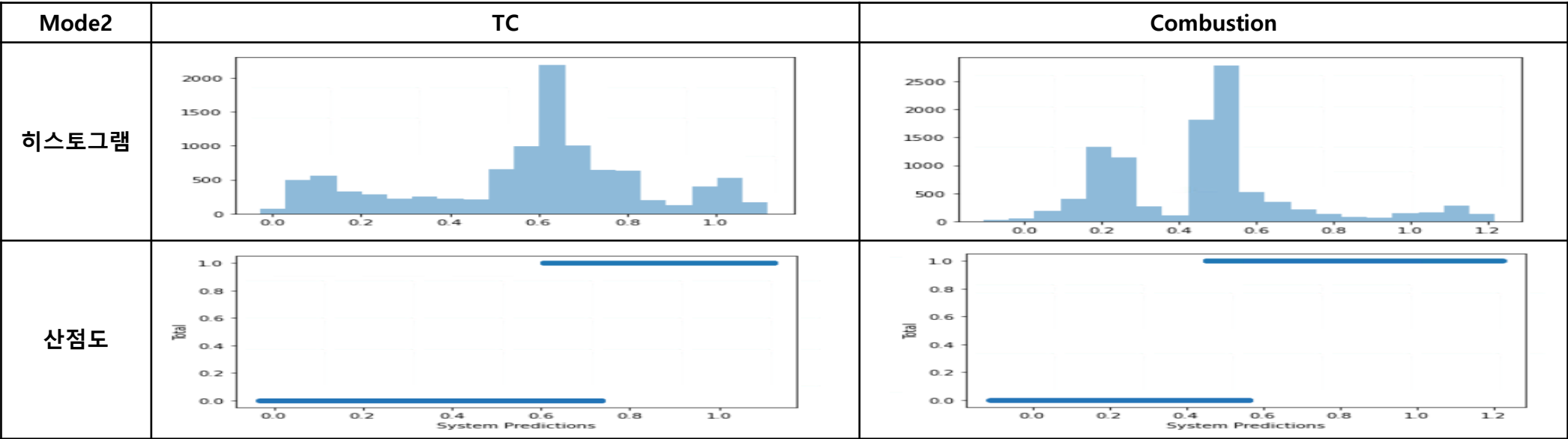
Train-mode2-Combustion





# 항목별 고장 진단 결과

## Train 데이터 전체



히스토그램 중앙부 가장 높은 도수를 가진 히스토그램의 계급을 만점(1, 정상),  
히스토그램의 좌측 군집에서 가장 높은 도수를 가진 히스토그램의 계급을 0점(0, 고장)으로 설정한 뒤  
그 범위에 속하지 않은 값들은 좌측이면 0, 우측이면 1로 설정.

0과 1사이를 linear scaling 하여 0과 1사이 값을 갖도록 하고 100을 곱해 최종 건강지수를 산출.



# Discussion

## 의의

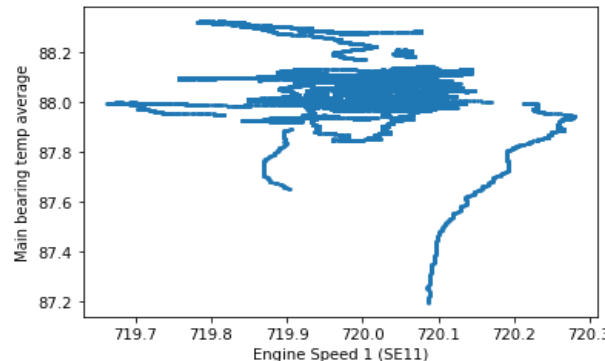
1. 수많은 센서 데이터와 내연기관의 운동에 대한 지식을 바탕으로 다양한 고장지표를 제안
2. 통계적인 지표를 활용해 점수를 부여
3. 비지도적 학습법이 혼재된 문제를 신경망의 회귀방식을 사용해 단순화하고, 분포를 분석

## 한계점

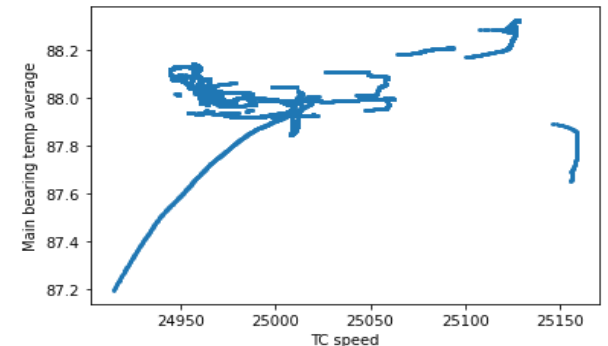
1. 주어진 데이터의 낮은 sampling rate(1Hz)로 인해 이상연소에 대한 정확한 판단이 힘들
2. 주어진 데이터로 학습을 시키면 정확히 전체 시스템이 고장인지 정상인지에 관해서만 판단할 수 있을 뿐, 어느 시스템이 문제가 발생하였는지는 확실하게 판단할 수 없음
3. Feature Selection method를 사용하지 못함

## 보다 나은 데이터 수집을 위한 제안

1. 가스 센서 - engine speed inlet과 BMEP로부터 추정된 CO, NOx의 양과 측정되는 양이 다르다면 엔진의 고장을 예측해볼 수 있음
2. 진동 센서 - 진동 센서를 사용하면 Fourier Transform 및 Wavelet Transform과 같은 주파수 영역 분석이 가능하여 다양한 고장 원인 파악 가능. 특히, 낮은 sampling rate로 높은 RPM의 엔진속도에서 나오는 신호들을 측정하기 어려운 한계 극복 가능.
3. 고장 원인의 라벨링 - 아무리 단순한 기계의, 한 가지의 고장이라도 원인을 분석하기 위해서 그만한 어려움이 있음. 적어도 몇 가지 뚜렷한 고장 모드를 나누어 함께 라벨링한다면 보다 효과적인 방식의 분석이 가능할 것으로 생각.
4. 데이터의 균형과 system 분류 - 엔진 스피드가 특정 구간(평형온도 인근)에 절대 다수가 분포해 엔진속도에 따라 상승하는 베어링의 온도를 적절히 반영하지 못함. 또한 적절한 관계가 나오더라도 구간이 너무 좁아(TC speed가 1% 정도, 온도가 1도 정도 변함) 이를 상관관계라고 보기 어려움. 이런 상황에서 특정 상황에만 맞추어 상관관계를 정의하면 베어링온도나 속도가 다를 때 이를 설명하지 못하는 현상이 발생.(좌측 하단 그림 참고)



엔진 스피드와 베어링 온도



TC 스피드와 베어링 온도

## 그외 방법들

1. GMM Clustering
2. Setting time과 감쇠 계수
3. Massive fusion



# References

- [1] Dimitrios T Hountalas, Prediction of marine diesel engine performance under fault conditions, *Applied Thermal Engineering*, Volume 20, Issue 18, 2000, Pages 1753-1783, ISSN 1359-4311, [https://doi.org/10.1016/S1359-4311\(00\)00006-5](https://doi.org/10.1016/S1359-4311(00)00006-5).
- [2] 정남용, 기계설계, 학진북스, 2020, p37
- [3] Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation, *Applied Thermal Engineering*, Volume 185, 2021, 116343, ISSN 1359-4311, <https://doi.org/10.1016/j.applthermaleng.2020.116343>.
- [4] Robert T. Balmer, *Modern Engineering Thermodynamics*, 2011
- [5] Robert T. Balmer, *Modern Engineering Thermodynamics*, 2011
- [6] Lee, Yang Min, Lee, Kwang-Young, Bae, Seung-Hyun, Jang, Hwi, and 이재기, “다채널 진동 센서를 이용한 선박 엔진의 진동 감지 및 고장 분류 시스템,” *정보처리학회논문지A*, vol. 17A, no. 2, pp. 81-92, Apr. 2010.
- [7] Xu Wang, Yanping Cai, Aihua Li, Wei Zhang, Yingjuan Yue, Anbo Ming, Intelligent fault diagnosis of diesel engine via adaptive VMD-Rihaczek distribution and graph regularized bi-directional NMF, *Measurement*, Volume 172, 2021, 108823, ISSN 0263-2241, <https://doi.org/10.1016/j.measurement.2020.108823>.
- [8] Educational and Psychological Measurement and Evaluation, 8th Edition. - Kenneth Hopkins, University of Colorado
- [9] L.H. Feng, J. Lu, The practical research on flood forecasting based on artificial neural networks, *Expert Syst. Appl.* 37 (2010) 2974-2977.
- [10] 김민수1906, context\_KKP, 제주감귤, DB분석가, TNT, 데이콘 경진대회 1등 솔루션, 위키북스, 2021
- [11] Yifeng Yang, Long Wei, Ying Hu, Yan Wu, Liangyun Hu, Shengdong Nie, Classification of Parkinson's disease based on multi-modal features and stacking ensemble learning, *Journal of Neuroscience Methods*, Volume 350, 2021, 109019, ISSN 0165-0270, <https://doi.org/10.1016/j.jneumeth.2020.109019>.
- [12] Su Hyun Youn, Taeyong Sim, Ahnryul Choi, Jinsung Song, Ki Young Shin, Il Kwon Lee, Hyun Mu Heo, Daeweon Lee, Joung Hwan Mun, Multi-class biological tissue classification based on a multi-classifier: Preliminary study of an automatic output power control for ultrasonic surgical units, *Computers in Biology and Medicine*, Volume 61, 2015, Pages 92-100, ISSN 0010-4825, <https://doi.org/10.1016/j.combiomed.2015.03.021>.
- [13] Joseba Castresana, Gorka Gabiña, Leopoldo Martin, Zigor Uriondo, Comparative performance and emissions assessments of a single-cylinder diesel engine using artificial neural network and thermodynamic simulation, *Applied Thermal Engineering*, Volume 185, 2021, 116343, ISSN 1359-4311, <https://doi.org/10.1016/j.applthermaleng.2020.116343>.
- [14] iwoong Lee, Inwoong Noh, Sang Ik Jeong, Yongho Lee, Sang Won Lee, Development of Real-time Diagnosis Framework for Angular Misalignment of Robot Spot-welding System Based on Machine Learning, *Procedia Manufacturing*, Volume 48, 2020, Pages 1009-1019, ISSN 2351-9789, <https://doi.org/10.1016/j.promfg.2020.05.140>.

# Q&A

**감사합니다!**