

---

# 목 차

1. 과제 배경 .....	1
2. 기존 사항 분석.....	1
3. 진행 방안 .....	2
3.1. 모델 구성 .....	2
3.2. 데이터 수집 .....	2
4. 예상 문제점 .....	2
5. 연구 방향 .....	3
6. 개발 일정 및 역할 분담.....	3
6.1. 개발 일정 .....	3
6.2. 역할 분담 .....	4
7. 참고 자료 .....	4

---

## 1. 과제 배경

일상 생활에서 기계를 제때 관리하지 않고 그대로 오래 쓰다 보면 고장 나는 경우가 많다. 이와 비슷하게 많은 기계를 사용하는 큰 산업에서도, 실제로 기계과부하시 기계장치에 고장이나 이상이 생기는 경우가 많다. 기업에서 이런 기계장치의 이상은 재산피해는 물론이고 심한 경우 인명피해까지 이르게 된다. 이런 피해를 줄이기 위해 최근 4차 산업혁명과 함께 많은 기업에서 다양한 AI 기술을 활용해 기계장치 이상진단에 관련된 연구를 진행하고 있다.

따라서 우리는 기업이 갖고 있는 시계열 데이터를 이용해 기계장치 이상을 진단하는 모델을 개발 및 고도화하면 기계장치 이상으로 인한 피해를 최소화 할 수 있을 것으로 판단되어 주제를 이와 같이 선정했다.

## 2. 기존 사항 분석

본 주제를 수행하기 앞서 인터넷을 통해 자료조사를 진행하였다. 그 중 빅데이터 분석에 의한 선박 엔진 고장 예측 알고리즘에 대한 논문을 볼 수 있었다[1]. 이 논문에서는 기존 대부분 선박에서 사용중인 AMS(Alarm Monitoring System)의 성능 한계로 인한 주요설비 고장을 문제점으로 말하며 컴퓨터 성능과 빅데이터 분석을 기반으로 한 머신러닝을 통해 고장에 대한 사전 예측 및 예방과 효율적인 관리를 할 수 있다고 한다. 더하여 해당 논문에서 제안한 알고리즘이 다양한 분야에 활용도가 높은 기술이라 말한다.

따라서 우리는 선박 엔진 이상을 진단하는 모델을 개발 및 고도화하면서 추가로 선박엔진의 이상이 발생하기 5분전의 데이터 패턴을 가중학습하여 이상치 판단에 더 관점을 두고 실제로 선박엔진에 문제가 발생한 데이터의 흐름 유사도를 Threshold값을 설정하여 문제가 일어날 확률을 계산한다. 이렇게 하면 이상이 생기는 시계열 데이터의 패턴을 파악할 수 있고, 본 과제에서도 머신러닝을 통해 고장에 대한 사전 예측 및 예방까지 할 수 있을 것이라 추측된다.

---

### 3. 진행 방안

#### 3.1. 모델 구성

우리는 해당 과제를 얻은 데이터를 기반으로 기계장치의 이상이 생겼는지 정상인지 구분하는 분류의 문제로 접근하고자 한다. 시작은 시계열 Tabular data를 XgBoost를 이용하여 Attribute들의 순위를 매기고 label에 영향이 큰 Attribute를 추출한다. 후에 전처리가 끝난 data를 K-fold cross validation 방법을 사용하여 Test set, Validation set, Train set으로 분류한다. Train Set으로 이상치 분류 모델을 생성하기 위해 Gradient Boosting Classification 방법을 사용한다. Gradient Boosting Algorithm은 Tabular format 데이터(엑셀형태와 같이 X-Y Grid로 되어있는 데이터)에 대한 예측에서 엄청난 성능을 보여주고, 머신러닝 알고리즘 중에서도 가장 예측 성능이 높다고 알려진 알고리즘이다. 만들어진 모델을 Validation set과의 MSE를 측정하여 평가한다. 최종적으로 만들어진 모델을 Train Set과 MSE를 측정하여 평가한다.

#### 3.2. 데이터 수집

이번 과제에서 사용한 데이터는 한 기업에서 실제로 사용중인 선박 엔진 중 특정 부분의 시계열 데이터를 사용했다. 이 데이터는 엔진에 시동이 걸리고 꺼질 때 까지 총 1524개의 Attribute에 대해서 총 16441개의 1초 간격의 연속적인 시계열 데이터다. Attribute에는 Analog 값과 값을 입력할 수 있는 Parameter값, Characteristic map값, Discrete값이 있다. 우리가 사용할 데이터는 센서에서 측정한 Analog 값을 사용한다.

### 4. 예상 문제점

현재 예상되는 문제점으로는 데이터 전처리 과정이다. 얻은 데이터의 Attribute가 많은만큼 특정 Attribute의 중요도나 Attribute간의 관계를 파악하기 어려울 것으로 예상된다. 이 문제점은 시계열 Tabular data를 XgBoost를 이용하거나 Attribute들의 순위를 매기고 label에 영향이 큰 Attribute추출하는 방식으로 어느정도 해결할 수 있을 것으로 예상하고 있다. 더하여 Attribute 간의 관계 파악은 데이터 상관관계 측정함수를 통해 heatmap을 그려봄으로써 파악할 예정이다.

둘째로, 데이터의 balance가 불균형한 상태이므로 데이터 불균형성을 해결하는 것도 중요하다. 데이터 불균형성을 해결하는 방법에는 Under Sampling이나 Over Sampling 등의 방법을 통해 해결해볼 예정이다.

## 5. 연구 방향

이번 과제에서는 선박 엔진의 특정 부분에 대한 센서 값을 이용하여 모델을 생성하였지만 어느정도 높은 성능이 나온다면 선박 엔진의 특정 부분만이 아닌 더 넓은 범위의 센서 값을 이용하면 엔진 전체의 이상치 진단을 파악할 수 있다. 더 나아가 선박엔진의 센서 값이 아닌 사용자가 입력하는 parameter값에 대한 최적 값을 제안할 수 도 있다. 최종적으로 선박엔진 뿐만 아니라 산업에서 사용되는 기계장치에 대한 이상치 진단 판단 모델 및 입력 parameter값의 최적 값을 제안하는 모델로 발전할 수 있다.

## 6. 개발 일정 및 역할 분담

### 6.1. 개발 일정

6월				7월				8월				9월			
6	13	20	27	4	11	18	25	1	8	15	22	5	12	19	26
데이터 분석															
데이터 전처리															
	기본 분류 모델 생성														
			모델 최적화 및 입력 데이터 수정												
						중간 보고서 작성									
							모델 최적화 및 입력 데이터 수정								
														최종 성능 측정	
														최종 보고서 작성 및 발표 준비	

---

## 6.2. 역할 분담

이름	역할
강금석	<ul style="list-style-type: none"><li>● 데이터 시각화</li><li>● 분류 모델 생성</li><li>● MSE를 이용하여 모델의 성능 측정 및 평가</li></ul>
김민수	<ul style="list-style-type: none"><li>● XgBoost를 이용하여 데이터의 Attribute 분석</li><li>● 데이터의 Balance 조정</li><li>● 모델 최적화</li></ul>
공통	<ul style="list-style-type: none"><li>● 필요 지식 공부</li><li>● 보고서 작성 및 발표 자료 준비</li></ul>

## 7. 참고 자료

- |   |
|---|
| <ul style="list-style-type: none"><li>[1] <a href="https://ieeexplore.ieee.org/document/9294450">https://ieeexplore.ieee.org/document/9294450</a></li><li>[2] <a href="https://blog.paperspace.com/gradient-boosting-for-classification">https://blog.paperspace.com/gradient-boosting-for-classification</a></li><li>[3] <a href="https://xgboost.readthedocs.io/en/stable/">https://xgboost.readthedocs.io/en/stable/</a></li><li>[4] <a href="https://3months.tistory.com/368">https://3months.tistory.com/368</a></li></ul> |
|---|