

---

## 목 차

1. 요구조건 및 제약 사항 분석에 대한 수정사항.....	1
1.1 요구 조건.....	1
1.2 기존 제약사항에 대한 수정사항.....	1
2. 설계 상세화 및 변경 내역.....	1
2.1. 데이터 전처리.....	1
2.2. 모델 학습.....	2
2.3. 모델 최적화.....	2
3. 갱신된 과제 추진 계획.....	3
4. 구성원 별 진척도.....	3
5. 보고 시점까지의 과제 수행 내용 및 중간 결과.....	4

---

## 1. 요구조건 및 제약 사항 분석에 대한 수정사항

### 1.1 요구 조건

최근 4차 산업혁명과 함께 많은 기업에서 다양한 AI 기술을 활용해 기계장치 이상진단에 관련된 연구를 진행하고 있다. 우리는 선박 엔진의 데이터를 사용해 선박 엔진의 부하를 분석하여 선박엔진의 이상을 진단하는 모델을 개발하려고 한다.

우리가 가진 데이터의 feature는 1524개로 매우 많으므로, feature selection 기법 중 하나인 XgBoost를 통해 엔진 부하에 가장 영향을 미치는 feature를 탐색 후 모델 학습에 사용할 feature를 선택한다. 선택한 feature를 사용해 전력부하를 예측하는 자기 회귀(Autoregressive)모델로 해당 과제를 풀고자 한다.

### 1.2 기존 제약사항에 대한 수정사항

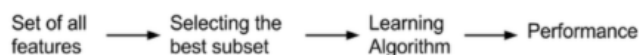
기존에 접근했던 방식은, 과부하 인지 아닌지 분류를 하는 모델이 기존 계획이었다. 하지만 실제 데이터에서의 값과 Classification을 통해 얻은 결과는 실제 데이터인 Analog 값과의 에러측정이 불가하다. 따라서 Classification 모델이 아닌, 실제로 우리가 사용하는 데이터의 부하 값과 예측한 부하 값 간의 MSE를 측정할 수 있는 자기회귀 (Autoregressive) 모델로 변경한다.

## 2. 설계 상세화 및 변경 내역

### 2.1. 데이터 전처리

- Filter Method를 활용해 feature간의 상관 관계를 알아낸 뒤, 어떤 feature가 상관관계수(영향력)가 높은지 알아내는 방법이다. correlation coefficient 방법을 사용해 correlation을 heatmap으로 표현하여 시각화 한다.

#### Filter Method



- 총 1524개의 Feature중 엔진의 센서를 통해 측정되는 Analog값만 이용을 한다. 그러나 해당 Feature도 총 875개이므로 Xgboost를 이용해서 모델 학습에 영향이 큰 Feature를 추출하여 사용할 Feature를 선택한다.

- 모델 학습에 사용되는 시계열 데이터는 총 16441개인데 입력 데이터의 label이 불균형하므로 Under Sampling이나 Over sampling을 이용해서 학습할 데이터의 balance를 맞춰준다.

---

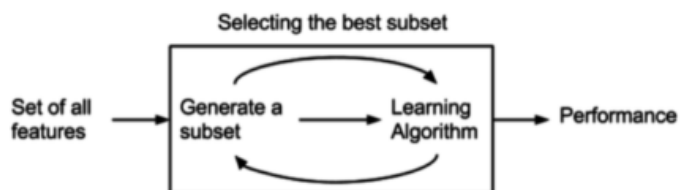
## 2.2. 모델 학습

- 특정 시간의 Random variable에 대해서 이전의 값이 이후의 값에 영향을 미치므로 자기상관(AR) 모델을 기본으로 한다.
- 시계열 데이터의 회귀분석에 사용되는 기법중 가장 많이 사용되는 기법중 하나인 AR과 MA를 통합한 ARMA 회귀 분석 모델을 사용한다.

## 2.3. 모델 최적화

- 회귀 분석 모델에서 에러를 측정하는데 보편적으로 사용되는 MSE 측정 방법을 사용한다.
- Wrapper Method 를 활용해 예측 정확도 측면에서 가장 좋은 성능을 보이는 Feature subset을 뽑아 낸다.

### Wrapper Method



- ARMA 회귀 분석 모델에서 사용되는 Hyper Parameter를 조절하여 모델의 성능을 높인다.

### 3. 갱신된 과제 추진 계획

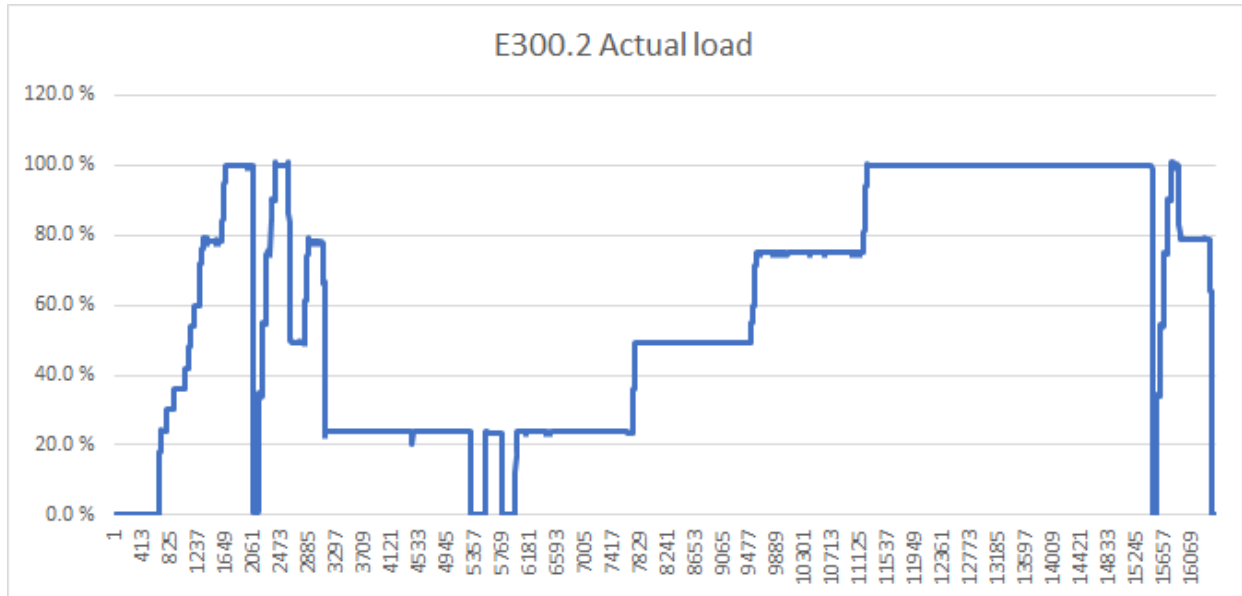
6월				7월				8월				9월			
6	13	20	27	4	11	18	25	1	8	15	22	5	12	19	26
데이터 분석															
데이터 전처리															
	기본 분류 모델 생성														
			모델 최적화 및 입력 데이터 수정												
						중간 보고서 작성									
							모델 최적화 및 입력 데이터 수정								
														최종 성능 측정	
														최종 보고서 작성 및 발표 준비	

### 4. 구성원 별 진척도

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

## 5. 보고 시점까지의 과제 수행 내용 및 중간 결과

### 5.1. 실제 예측해야 할 label의 그래프



### 5.2 Feature 그래프

