

# A study on Data Preprocessing for Developing Remaining Useful Life Predictions based on Stochastic Degradation Models Using Air Craft Engine Data

Yeon Ah Yoon\* · Jin Hyeong Jung\* · Jun Hyoung Lim\*\* · Tai-Woo Chang\*\*\* · Yong Soo Kim\*\*\*<sup>†</sup>

\*Department of Industrial and Management, Kyonggi University Graduate School

\*\*Intelligent System Engineering Division, Hancor MDS

\*\*\*Department of Industrial and Management, Kyonggi University

## 항공엔진 열화데이터 기반 잔여수명 예측력 향상을 위한 데이터 전처리 방법 연구

윤연아\* · 정진형\* · 임준형\*\* · 장태우\*\*\* · 김용수\*\*\*<sup>†</sup>

\*경기대학교 일반대학원 산업경영공학과

\*\*한컴MDS Intelligent System Engineering 사업본부

\*\*\*경기대학교 산업경영공학과

Recently, a study of prognosis and health management (PHM) was conducted to diagnose failure and predict the life of air craft engine parts using sensor data. PHM is a framework that provides individualized solutions for managing system health. This study predicted the remaining useful life (RUL) of aeroengine using degradation data collected by sensors provided by the IEEE 2008 PHM Conference Challenge. There are 218 engine sensor data that has initial wear and production deviations. It was difficult to determine the characteristics of the engine parts since the system and domain-specific information was not provided. Each engine has a different cycle, making it difficult to use time series models. Therefore, this analysis was performed using machine learning algorithms rather than statistical time series models. The machine learning algorithms used were a random forest, gradient boost tree analysis and XG boost. A sliding window was applied to develop RUL predictions. We compared model performance before and after applying the sliding window, and proposed a data preprocessing method to develop RUL predictions. The model was evaluated by R-square scores and root mean squares error (RMSE). It was shown that the XG boost model of the random split method using the sliding window preprocessing approach has the best predictive performance.

**Keywords :** Remaining Useful Life, Degradation, Preprocessing, Machine Learning, Sliding Window

## 1. 연구 개요

최근 센서 기술의 발전으로 센서 데이터의 활용도가 늘어나는 추세이며 여러 기능이 결합된 복잡한 시스템에서 고장이 발생하기 전 유지보수를 위한 예방보전의 중요성이 대두되고 있다. 또한 일반적인 제조, 항공 및 국방 등 많은 분야에서 고장예지 및 건전성관리(prognosis and health management, PHM)에 대해 활발히 논의되고 있으며 일부 분야에서는 적용 중에 있다[2]. 고장예지 및 건전성 관리 기술은 크게 물리 기반 접근법, 데이터주도 접근법, 융합 접근법으로 구분된다[19]. 본 논문에서는 데이터주도 접근법 중 머신러닝 기법을 통해 항공 엔진의 잔여 수명을 예측하는 모델을 구축하고 모델의 예측 성능을 향상시키기 위한 데이터 전처리 기법을 제안하고자 한다.

사용한 데이터는 2008 PHM Conference Challenge에서 배포한 센서 기반의 수집 데이터로 총 218개의 항공엔진 정보를 다루고 있다. 각 엔진은 알려지지 않은 초기 마모 및 제작 편차가 존재하며 시스템 및 도메인별 정보가 별도로 제공되지 않았다. 일반적인 수명 추정을 위한 열화데이터 분석은 시험대상 부품 특성에 따라 여러 단계별 통계적 추정과 예측을 실시하여 열화모델을 선정한다[20]. 그러나 본 연구에서는 부품 특성을 파악하기에 어려움이 존재하므로 머신러닝 모델을 활용하여 측정값들을 통해 잔여수명을 예측하고자 한다.

일반적인 시계열 모델을 사용하기에 각 엔진별 고장이 발생하는 시점이 다르므로 시퀀스 길이 및 모델 파라미터를 지정하기에 한계가 존재한다. 따라서 본 연구에서는 시계열 모델이 아닌 다른 정적모델(static model)을 통해 잔여 수명을 예측하고자 하였다. 각 모델별 분할 방법을 다르게 지정하였으며 예측 성능 향상을 위해 데이터 전처리 과정에 슬라이딩 윈도우(sliding window)기법[7]을 제안하였다.

본 논문에서는 제 2장에서 관련된 문헌들을 소개하고 제 3장에서 연구 프로세스를 소개한다. 제 4장을 통해 실험결과를 보이고 마지막으로 제 5장에서 결론 및 추후 연구 과제를 제시한다.

## 2. 관련 문헌 연구

관련 문헌 연구는 크게 세 가지로 나누어 수행하였다. 첫 번째로, 본 연구에서 사용한 동일한 데이터를 통해 항공 엔진의 잔여수명을 예측한 방법론에 대한 문헌연구를 수행하였다. Heimes[11]는 해당 데이터를 다변량 시계열 데이터 측면에서 모든 데이터를 활용한 잔여수명 예측을 위해 MLP(multi-layer perceptron) 및 RNN(recurrent neural network) 모델을 구축하였다. Son et al.[15]은 주성분 분석을 수행하고 성능저하지표를 구축한 후 확률론적 예측 모델

링을 위해 위너 프로세스(wiener process)를 활용하였다. 이외에도 효과기반 예측을 위한 열화 매개변수를 식별하고, 칼만필터(Kalman filter)와 인공신경망의 앙상블 모델을 구축하는 등 다양한 방법론을 통해 항공 엔진의 잔여 수명을 예측하였다[9, 17, 21].

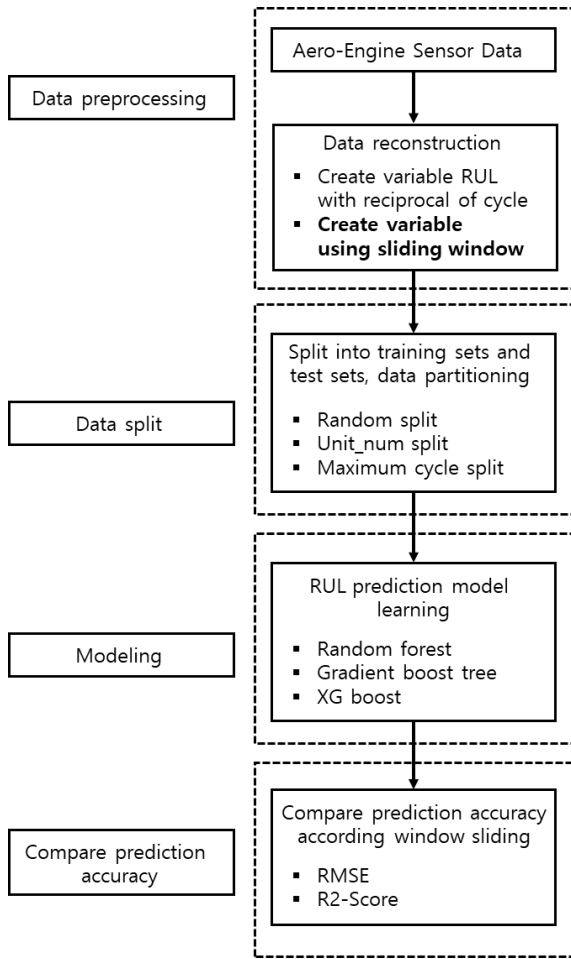
두 번째로 열화데이터 분석 시 머신러닝의 활용성을 파악하기 위한 관련 문헌 연구를 수행하였다. Choi et al.[8]은 표면조직복제법으로 고온 부품의 열화 상태를 평가하는 방법에 머신러닝을 적용하여 기존보다 객관적인 열화 평가 시스템을 제안하였다. Jung and Kim[12]은 철도차량 부품 관리의 효율적인 유지보수 시스템 구축을 위해 머신러닝 기반의 철도차량 고장진단 모델을 개발하였다. Bomtepi et al.[3]은 예측모델을 공식화하고 시계열 데이터 처리를 위한 학습 기법에 초점을 맞추어 예측 수행 시 사용될 머신러닝 기법을 제안하였다.

마지막으로 본 논문에서 제안하는 데이터 전처리 방법인 슬라이딩 윈도우 기법에 관한 관련 문헌 연구를 수행하였다. Lee and Oh[16]는 슬라이딩 윈도우 기법을 활용하여 시간적 단계에 따라 신용평점을 예측하는 모델을 구축하였다. Seo et al.[18]은 윈도우 단위의 전처리를 통해 연속된 스트림 데이터에 대한 다차원 속성의 분류 기법을 제안하였다. Ann[1]은 휠로더(wheel loader) 엔진의 센서 데이터를 활용하여 장비의 상태 분류를 예측하는 새로운 CNN 모델을 제안하였다. 모델 구축 시 센서 데이터의 시그널 특성을 학습시키기 위해 슬라이딩 윈도우 기법을 활용하였으며 이동 구간으로 추출된 센서데이터를 이미지 형태로 인식하였다.

열화데이터 분석을 위한 다양한 방법론이 연구되어 왔으며 슬라이딩 윈도우 기법 또한 다방면으로 연구가 수행되어 왔다. 그러나 정적 모델을 통해 시계열 데이터의 잔여수명을 예측한 연구는 미흡한 상황이다. 따라서 본 연구에서는 시계열 모델이 아닌 머신러닝 기법을 통해 잔여수명을 예측하고, 예측 성능을 향상시키고자 데이터 전처리 과정에 슬라이딩 윈도우 기법을 적용하는 방법을 제안하고자 한다.

## 3. 연구 프로세스

본 논문에서 제안하는 연구 프로세스는 <Figure 1>과 같다. 먼저 모델 구축을 위해 종속변수를 산출하고 독립변수를 정의하였다. 그 후 전체 218개의 엔진의 잔여수명 예측을 위한 모델을 구축하였다. 모델 구축 시 훈련데이터와 시험데이터를 세 가지 방법으로 분할하였으며 모델링 및 잔여수명 예측은 Dataiku Data Science Studio (DSS)를 활용하였다. 슬라이딩 윈도우 기법을 적용하여 각 모델의 잔여수명 예측 성능 변동을 확인하였다. 엔진별 수명길이가

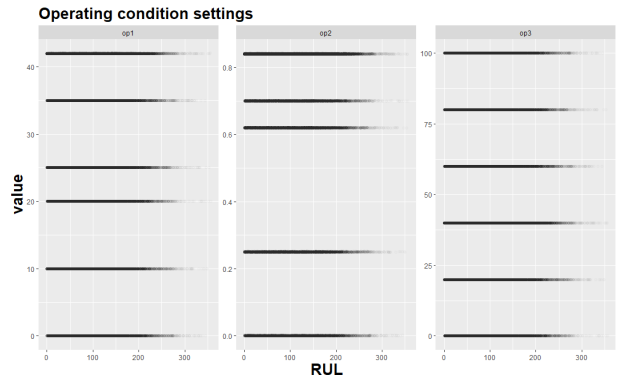


<Figure 1> Research Process

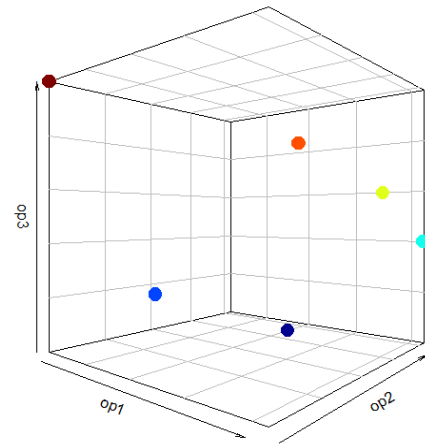
상이하므로 수명길이에 따라 나누어 모델링할 경우 예측 성능에 변동이 존재할 수 있다. 따라서 운행횟수에 따라 클러스터링을 수행하여 단기, 중기, 장기로 나눈 후 전체 데이터 모델링 과정과 동일하게 분석을 진행하였다.

먼저 전체 데이터의 변수들을 재정의하였다. 본 논문에서 사용된 데이터는 서로 다른 218개의 항공기 엔진에 대한 센서 데이터로 각 엔진은 알려지지 않은 초기 마모 및 제작 편차가 존재한다. 수명이 서로 다른 218개의 열화데이터 셋 (unit)으로 구성되어 있으며 엔진별 설정값(operation, op) 3개와 측정값(measurement, mea) 21개가 존재한다. 데이터의 변수는 엔진 ID(unit), 엔진 운행횟수(cycle), 설정값 3개(op1-3), 측정값 21개(meal-21)로 총 26개로 구성되어 있다.

각 엔진별 운행횟수가 다르며 마지막 운행시점에서 시스템 고장이 발생하였다는 데이터의 설명 하에 엔진의 운행횟수를 수명 길이로 정의하였다. 데이터셋 내에 종속변수가 존재하지 않으므로 엔진별 최대 운행횟수와 수행된 운행횟수의 차를 잔여수명(RUL)으로 정의하여 새로운 종속변수로 생성하였다. RUL 산출방식은 아래 수식과 같으며  $l$ 은 엔진 ID,  $t_{max}$ 는 엔진별 최대 운행횟수,



<Figure 2> op1, op2, op3 by RUL



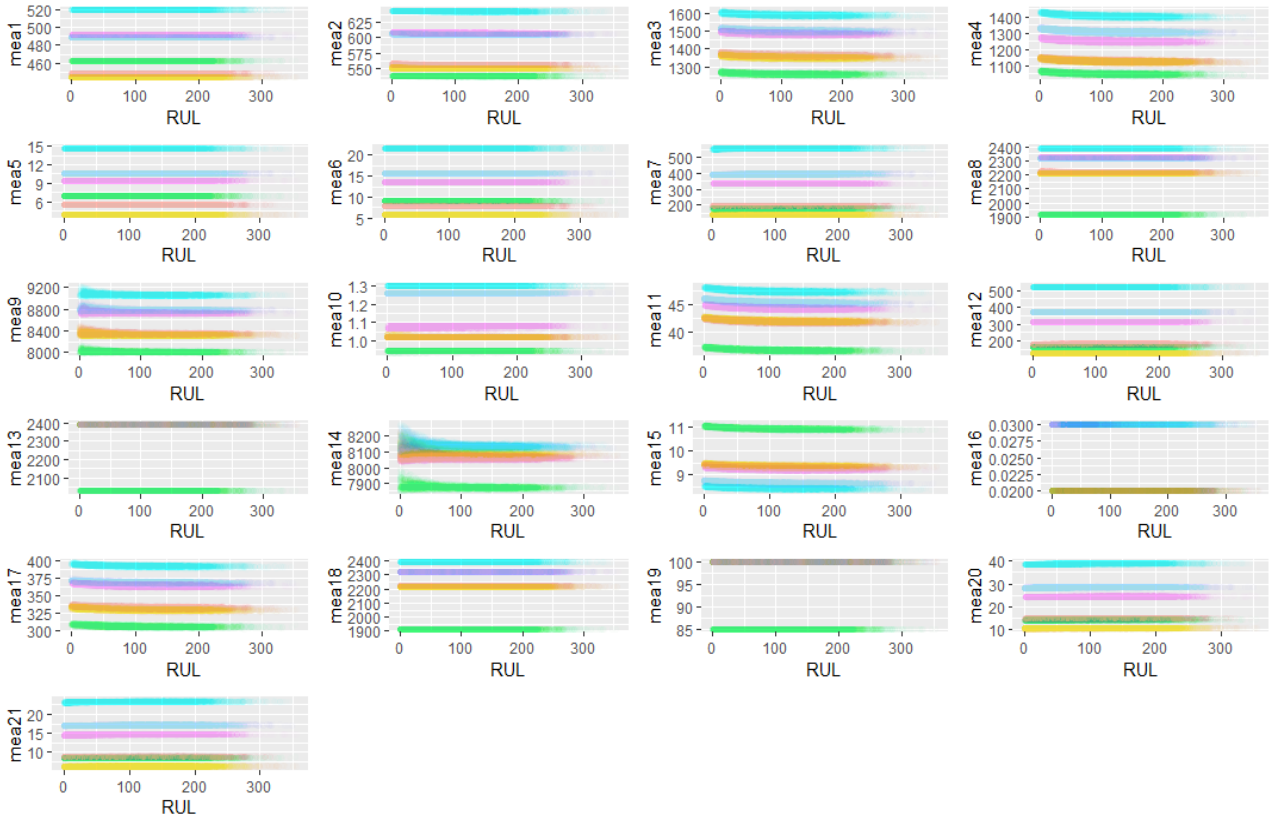
<Figure 3> Three Operation Setting Variables

$t_c$ 은 엔진별 수행된 운행횟수를 의미한다.

$$RUL^{(l)} = t_{max} - t_c$$

<Figure 2>는 각 RUL에 따른 엔진별 설정값들(op)의 추세를 시각화한 결과이다. 이를 통해 설정값들이 5~6개의 군집을 이루고 있으므로 설정값 변수 3개가 각기 다른 작동모드를 의미함을 추정할 수 있다. 따라서 설정값들의 k-means 클러스터링을 수행하였으며 결과는 <Figure 3>과 같다. 이를 통해 총 6개의 작동모드가 존재함을 알 수 있다. 이후 6개의 작동모드별 RUL에 따른 측정값의 경향성을 확인하기 위해 도식화한 결과는 <Figure 4>와 같다. <Figure 4>를 통해 특정 측정값이 RUL에 따른 경향성이 존재하며 6개의 작동모드로 구분됨을 확인하였다.

모델링 수행 시 6개의 작동모드를 설명하는 op\_ID를 추가하고[15] 종속변수를 RUL로 설정하였다. 따라서 사용된 독립변수는 op\_ID와 meal-21로 총 22개의 변수이다. 사용한 머신러닝 기법은 총 3가지로 랜덤포레스트(random forest), 그래디언트 부스트(gradient boost tree), XG 부스트(XG boost)이다. 각 모델들 학습 시 데이터에 맞추어 Dataiku Data Science Studio(DSS) 툴을 통해 파라미터 최적화가 수행되었다.



<Figure 4> Measurements(meas01~21) by RUL

랜덤포레스트는 Breiman에 의해 제안된 앙상블(ensemble) 기반 모형으로 의사결정나무 모형에 배깅(Bagging)의 기본 원리와 임의성을 더한 형태이다[4]. 그래디언트 부스트는 Friedman에 의해 제안된 방법으로 의사결정나무에 비해 여러 분류자의 예측을 종합하여 분류의 정확성을 향상하는 앙상블 기법 중에 하나이다[10]. XG 부스트는 Chen and Guestrin이 제시하였으며 의사결정나무의 부스팅 기법을 병렬처리를 통해 학습속도를 개선하여 우수성을 입증하였다[6].

센서 데이터의 경우 시그널 특성이 존재하므로 한 시점이 아닌 패턴 변화를 감지해야한다[14]. 정적인 머신러닝 기법을 통해 데이터의 패턴 변화 및 시계열성을 다루기에 어려움이 존재하므로 데이터 내에 시계열성을 포함시키고자 윈도우 기법을 통한 데이터 전처리 방법을 제안하였다.

윈도우 기법 중 슬라이딩 윈도우 기법은 데이터를 시간의 흐름에 따라 순차적으로 윈도우 크기만큼의 새로운 관측치를 정의하는 것이다[5, 13]. 윈도우 크기는 3으로 설정하였으며 측정값 변수들의 하향 및 상향 추세를 반영하고자 최소값과 최대값을 추출하여 윈도우 변수를 생성하였다. 윈도우 변수 산출방식은 다음 수식과 같다.

$$X_{\min} = \min(x_i^\alpha, x_{i+1}^\alpha, \dots, x_{i+(s-1)}^\alpha)$$

$$y_{\min, i}^\alpha = \begin{cases} \min(X_{\min}, y_{i-1}), & X_{\min} < x_{i+s}^\alpha \\ \min(y_{i-1}, x_{i+s}^\alpha), & X_{\min} > x_{i+s}^\alpha \end{cases}$$

$$X_{\max} = \max(x_i^\alpha, x_{i+1}^\alpha, \dots, x_{i+(s-1)}^\alpha)$$

$$y_{\max, i}^\alpha = \begin{cases} \max(X_{\max}, y_{i-1}), & X_{\max} > x_{i+s}^\alpha \\ \max(y_{i-1}, x_{i+s}^\alpha), & X_{\max} < x_{i+s}^\alpha \end{cases}$$

$x_i^\alpha$ 는  $\alpha$ 의 각 측정값 변수(meas01-21)의  $i$ 번째 운행횟수에서의 기존 측정값을 의미한다.  $s$ 는 슬라이딩 윈도우 크기를 의미하며  $y_i^\alpha$ 는 생성되는 윈도우 변수를 의미한다. 따라서 슬라이딩 윈도우 기법을 적용한 후 사용된 독립변수는 추가된 최소, 최대 윈도우 변수 42개와 기존 변수들 21개를 합친 63개의 측정값 변수와 명목형 독립변수인 op\_ID로 총 64개로 구성된다.

일반적인 경우 훈련데이터와 시험데이터를 분할할 시 랜덤하게 분할하나, 슬라이딩 윈도우 변수의 시계열성이 손상될 우려가 있다. 따라서 본 연구에서는 랜덤 분할, 엔진별 분할, 최대 사이클별 분할로 방법을 나누어 모델링하였으며 각 분할 방법과 슬라이딩 윈도우 기법에 따른 예측 성능의 변동을 확인하였다.

각 엔진별 수명길이가 다르므로 수명길이에 따라 나누어 모델링할 경우 예측 성능에 변동이 존재할 수 있다. 따라서 수명길이를 기준으로 클러스터링을 통해 단기, 중기, 장기 수명을 지닌 엔진으로 나누어 데이터셋을 구축하였다. 모델링 과정은 전체 데이터 모델링 과정과 동일하게 진행하였으며 예측 성능의 변동을 확인하였다.

성능이 크게 향상되었으나, 다른 분할 방법에서는 큰 차이를 보이지 않았다. 이는 엔진별 수명길이의 차이로 인한 현상으로 사료되었으며 수명길이에 따라 엔진을 나누어 모델링을 진행할 시 전체 데이터를 통해 모델링하는 것에 비해 예측 성능이 향상될 것이라 판단하였다. 클러스터링을 통해 단기, 중기, 장기로 수명길이별 군집을 형성하였으며 각 군집별 모델을 구축하였다.

### 4. 실험분석 및 결과

#### 4.1 전체 데이터를 통한 잔여수명 예측

전체 218개의 엔진에 대해 앞서 언급한 슬라이딩 윈도우 기법 사용 여부 및 분할방식별로 세 가지 모델을 통한 예측 성능 결과는 <Table 1>과 같다. 슬라이딩 윈도우 기법을 사용하지 않은 경우, 분할 방법에 따른 예측 성능은 엔진별 분할 시 예측력이 가장 높았으나 나머지 분할 방법과 큰 차이를 보이지 않았다. 그러나 슬라이딩 윈도우 기법을 적용한 경우, 랜덤 분할 시 XG 부스트 모델의 성능이  $R^2$ 은 0.8368, RMSE는 27.718로 가장 높음을 확인하였다.

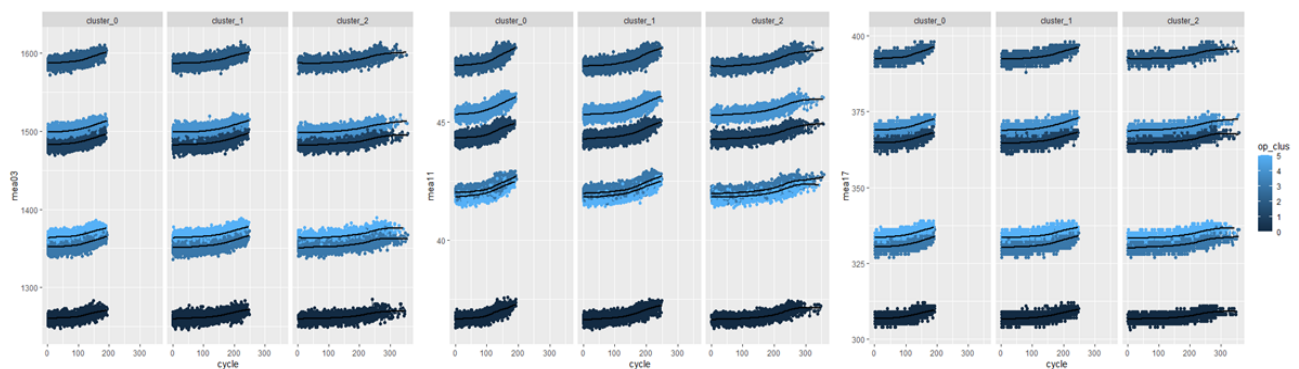
랜덤 분할 시 슬라이딩 윈도우 기법을 적용하여 예측

#### 4.2 수명길이 군집별 잔여수명 예측

운행횟수를 기준으로 클러스터링을 수행하여 수명길이가 단기(cluster\_0), 중기(cluster\_1), 장기(cluster\_2)인 엔진으로 군집을 형성하였다. 그 결과 단기 82개, 중기 98개, 장기 38개의 엔진으로 구성되었다. 군집별 잔여수명예측을 수행할 시 같은 추세를 보여야하므로 이를 확인하기 위해 데이터 시각화를 진행하였다. <Figure 5>는 대표적으로 mea03, mea11, mea17에 대한 추세 그림이다. 이를 통해 수명길이에 따른 차이가 존재할 뿐 수명 군집별로 비슷한 추세를 보임을 확인하였다. 따라서 수명길이별로 나누어 잔여수명을 예측하였으며 각 수명 군집별로 전체 데이터를 학습한 방법과 동일하게 모델링하였다. 모델링 수행결과는 <Table 2>와 같다.

<Table 1> Result of Modeling on Whole Data

Split way	Random		unit_ID		Maximum cycle	
	X	O	X	O	X	O
Using window						
Modeling	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)
Random forest	0.6435 (40.963)	0.8228 (28.925)	0.6601 (38.452)	0.5958 (41.929)	0.5971 (45.412)	0.5389 (48.578)
Gradient Boosted Trees	0.6091 (42.895)	0.7047 (37.283)	0.6263 (40.3203)	0.6331 (39.951)	0.5586 (47.503)	0.5510 (47.935)
XG Boost	0.6306 (41.700)	<b>0.8368</b> <b>(27.718)</b>	0.6463 (39.223)	0.6719 (37.778)	0.5818 (46.2633)	0.5737 (46.7058)



<Figure 5> Trend of mea03, mea11, and mea17

<Table 2> Result of Modeling on Groups of Unit

Short cycle units						
Split way	Random		unit_ID		Maximum cycle	
Using window	X	O	X	O	X	O
Modeling	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)
Random forest	0.6615 (28.934)	0.7938 (22.583)	0.7026 (26.619)	0.6459 (29.047)	0.6270 (29.556)	0.6354 (29.222)
Gradient Boosted Trees	0.6392 (29.876)	0.7316 (25.764)	0.6574 (27.812)	0.6789 (27.661)	0.6114 (30.168)	0.6794 (27.401)
XG Boost	0.6394 (29.867)	<b>0.8388</b> <b>(19.965)</b>	0.6825 (27.508)	0.7115 (26.219)	0.6160 (29.989)	0.7501 (24.191)
Medium cycle units						
Split way	Random		unit_ID		Maximum cycle	
Using window	X	O	X	O	X	O
Modeling	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)
Random forest	0.7146 (34.325)	0.8203 (27.313)	0.7168 (34.264)	0.7393 (32.877)	0.6465 (39.660)	0.7355 (34.303)
Gradient Boosted Trees	0.6847 (36.183)	0.7656 (31.197)	0.6913 (35.773)	0.7358 (33.096)	0.6231 (40.948)	0.7002 (36.519)
XG Boost	0.711 (35.071)	<b>0.8472</b> <b>(25.182)</b>	0.7123 (34.538)	0.7789 (30.276)	0.6383 (40.115)	0.7719 (31.854)
Long cycle units						
Split way	Random		unit_ID		Maximum cycle	
Using window	X	O	X	O	X	O
Modeling	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)	$R^2$ (RMSE)
Random forest	0.7248 (42.614)	0.8271 (33.781)	0.7058 (43.818)	0.7500 (40.397)	0.7098 (44.771)	0.6951 (45.885)
Gradient Boosted Trees	0.6942 (44.927)	0.7905 (37.185)	0.6740 (46.132)	0.7086 (43.613)	0.6707 (47.693)	0.6873 (46.471)
XG Boost	0.7048 (44.136)	<b>0.8487</b> <b>(31.599)</b>	0.6752 (46.047)	0.7499 (40.401)	0.6752 (47.364)	0.7488 (41.648)

수명군집별 예측 성능이 가장 우수한 모델은 슬라이딩 윈도우 기법을 적용한 랜덤 분할 방식의 XG부스트 모델로 동일하였다. 윈도우 기법 적용 전, 분할방법에 따른 예측력 변동을 확인하였을 때 엔진별 기준으로 분할하는 방법이 우수하며 큰 변동을 보이지는 않았다. 그러나 윈도우 기법을 적용할 시 전반적으로 예측 성능이 크게 향상됨을 확인하였다.

### 5. 결론 및 추후 연구과제

본 연구는 열화 특성 파악에 어려움이 존재하는 항공 엔진의 잔여 수명을 예측하기 위해 머신러닝 기법을 활용하였다. 엔진별로 시퀀스 길이가 다르므로 시계열 모델이 아닌 정적인 머신러닝 기법을 활용하였다. 데이터에 임의로 시계열적인 특성을 부가하여 데이터 전처리

수행 시 슬라이딩 윈도우 기법을 사용하였으며, 다양한 분할방법을 통해 모델링 한 후 예측 성능을 비교하였다.

전체 데이터에 대한 모델링 및 엔진수명에 따른 각 군집별 모델링을 수행한 결과, 군집별로 잔여 수명을 예측하는 것이 비교적 우수한 성능을 보이는 것을 확인하였다. 분할방법에 따른 예측 성능의 변동은 미미하였으며, 슬라이딩 윈도우 기법 적용 시 전반적으로 예측 성능이 향상되었다. 실험결과 슬라이딩 윈도우 기법을 적용한 랜덤 분할 방식의 XG 부스트 모델이 가장 우수한 예측 성능을 보임을 확인하였다.

정적 모델링 수행 시, 데이터의 시계열성이 손실될 수 있으나 랜덤 분할 방식의 예측성능이 가장 우수함을 확인하였다. 전체 데이터 학습의 경우, 데이터 분할 시 수명길이 편향되어 학습될 가능성이 존재하므로 엔진별 및 최대 수명길이별 분할이 랜덤 분할 방식에 비해 낮은 성능을 보인다고 사료된다. 그러나 편향을 고려하여 수명길이별

단기, 중기, 장기로 데이터셋을 나누어 학습을 진행한 결과 또한 랜덤분할 방식의 예측성능이 가장 우수하였다. 이는 초기 마모 편차에 의한 열화추세 및 열화발생시점의 차이에 의한 현상이라고 사료된다.

본 논문에서는 슬라이딩 윈도우 기법을 활용한 데이터 전처리 과정이 예측 성능에 어떠한 영향을 미치는지 확인하고자 하였으며, 수명길이와 관계없이 동일한 윈도우 사이즈를 설정하였다. 그러나 윈도우 사이즈에 따라 도출되는 관측치가 다르므로 수명길이에 따라 다르게 설정할 시 보다 우수한 예측모델을 구축할 수 있을 것으로 사료된다. 또한, 본 연구에서는 시계열 모델 사용에 한계가 존재할 시 데이터 변환을 통한 예측 성능 변동을 확인하고자 시계열 모형을 고려하지 않았으나 추후 시계열 모델 활용할 시 더욱 높은 예측 성능을 보일 수 있을 것이라 기대된다.

본 연구에서 사용한 항공 엔진 데이터만이 아닌 유사한 특성을 지닌 데이터에도 적용할 수 있을 것이다 기대된다. 본 연구를 통해 시스템 및 도메인 정보를 알 수 없으며 시계열 모델을 활용하기 어려운 경우, 제안한 전처리 방법을 활용하여 예측 성능 향상을 도모할 수 있다.

## Acknowledgement

This work was supported by Kyonggi University's Graduate Research Assistantship 2020.

## References

- [1] Ann, G., Yoo, J.H., Lee, S.H., and Kim, S.B., Explainable convolution neural networks for multi-sensor data, *Journal of the Korean Institute of Industrial Engineers*, 2019, Vol. 45, No. 2, pp. 146-153.
- [2] Baik, J., AI techniques for prognostics and health management, *Journal of Applied Reliability*, 2019, Vol. 19, No. 3, pp. 243-255.
- [3] Bontempi, G., Taieb, S.B. and Le Borgne, Y.A., Machine learning strategies for time series forecasting, *European Business Intelligence Summer School*, Springer, Berlin, Heidelberg, 2012, pp. 62-77.
- [4] Breiman, L., Random forests, *Machine Learning*, 2001, Vol. 45, No. 1, pp. 5-32.
- [5] Chang, J.-H., A fuzzy window mechanism for information differentiation in mining data streams, *Journal of the Korea Academia-Industrial Cooperation Society*, 2011, Vol. 12, No. 9, pp. 4183-4191.
- [6] Chen, T. and Guestrin, C., Xgboost : A scalable tree boosting system, *Proceedings of the 22nd acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.
- [7] Chiu, S.C., Li, H.F., Huang, J.L., and You, H.H., Incremental mining of closed inter-transaction itemsets over data stream sliding window, *Journal of Information Science*, 2011, Vol. 37, No. 2, pp. 208-220.
- [8] Choi, W., Chang, S., Lee, S., Kang, H., Bang, M., and Bae, Y., Development of degradation evaluation SW for high temperature component using machine learning approach, *The Korean Society of Mechanical Engineers*, 2020, Vol. 44, No. 1, pp. 57-62.
- [9] Coble, J.B. and Hines, J.W., Prognostic algorithm categorization with PHM challenge application, *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1-6.
- [10] Friedman, J.H., Greedy function approximation : a gradient boosting machine, *Annals of Statistics*, 2001, pp. 1189-1232.
- [11] Heimes, F.O., Recurrent Neural Networks for Remaining Useful Life Estimation, *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1-6.
- [12] Jung, H. and Kim, J.-W., A machine learning approach for mechanical motor fault diagnosis, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 1, pp. 57-64.
- [13] Koc, C.K., Analysis of sliding window techniques for exponentiation, *Computers and Mathematics with Applications*, 1995, Vol. 30, No. 10, pp. 17-24.
- [14] Laguna, J.O., Olaya, A.G., and Borrajo, D., A dynamic sliding window approach for activity recognition, *International Conference on User Modeling, Adaptation, and Personalization*, pringer, Berlin, Heidelberg, 2011, pp. 219-230.
- [15] Le Son, K., Fouladirad, M., Barros, A., Levart, E., and Lung, B., Remaining useful life estimation based on stochastic deterioration models : A comparative study, *Reliability Engineering and System Safety*, 2013, Vol. 112, pp. 165-175.
- [16] Lee, H.S. and Oh, S.H., A Study on the development of time series forecasting model for corporate credit risk using machine learning, *The Korean Society of Management Information Systems*, Seoul, Korea, 2019, pp. 396-405.
- [17] Peel, L., Data driven prognostics using a kalman filter

ensemble of neural network models, *2008 International Conference on Prognostics And Health Management*, IEEE, 2008, pp. 1-6.

- [18] Seo, S., Kang, J., Nam, K.W., and Ryu, K.H., A sliding window-based multivariate stream data classification, *The Korean Institute of Information Scientists and Engineers*, 2006, Vol. 33, No. 2, pp. 163-174.
- [19] Sim, H.S., Kang, J.-G., and Kim, Y.S., A review on prognostics and health management : 2013~2018, *Journal of Applied Reliability*, 2019, Vol. 19, No. 1, pp. 68-84.
- [20] Song, H.S., Seo, Y.K., Jung, D.H., and Park, B.H., A case study of degradation analysis for the passenger vehicles shock absorber, *Journal of Applied Reliability*,

2017, Vol. 17, No. 3, pp. 181-187.

- [21] Wang, T., Yu, J., Siegel, D., and Lee, J., A similarity-based prognostics approach for remaining useful life estimation of engineered systems, *2008 International Conference on Prognostics and Health Management*, IEEE, 2008, pp. 1-6.

#### ORCID

Yeon Ah Yoon | <http://orcid.org/0000-0001-5656-8402>  
Jin Hyeong Jung | <http://orcid.org/0000-0001-9414-4571>  
Jun Hyoung Lim | <http://orcid.org/0000-0002-6422-3700>  
Tai-Woo Chang | <http://orcid.org/0000-0003-3563-206X>  
Yong Soo Kim | <http://orcid.org/0000-0003-3362-4496>