

# 무인항공기의 불균형 수리부속에 대한 수요예측

최정호<sup>1</sup>, 유지훈<sup>2</sup>, 김재동<sup>2\*</sup>

<sup>1</sup>광운대학교 방위사업학과(박사과정)/육군 소령, <sup>2</sup>한국국방연구원

## Demand Forecast for imbalanced spare parts of UAV

Jung-Ho Choi<sup>1</sup>, Ji-Hoon Yu<sup>2</sup>, Jae-Dong Kim<sup>2\*</sup>

<sup>1</sup>Department of Defense Business, KwangWoon University(Phd Candidate)

<sup>2</sup>Korea Institute For Defense Analyses

**요약** 최근 전략무기체계인 무인항공기에 대한 세계적 관심이 증가하고 있다. 이러한 세계적 추세에 맞춰 한국군도 무인항공기 개발을 통한 전투준비태세 유지에 힘쓰고 있다. 무인항공기를 효과적으로 운용하기 위해서는 수리부속의 보급 및 정비가 필수적이다. 이를 위해서는 수리부속 수요를 정확하게 예측하는 것이 중요하다. 그동안 한국군은 장비정비 정보체계의 데이터를 활용하여 몇 가지 시계열 방법을 기반으로 수리부속 수요를 예측해 왔으나, 수리부속 수요예측의 정확성을 높이기 위한 개선된 예측모델이 필요하다. 본 연구에서는 오버샘플링과 언더샘플링을 비교 분석하여 수리부속 수요예측의 정확도를 높일 수 있는 모델을 제안하며, 데이터 분포의 일관성 강화를 위해 전처리 기술을 적용하여 정확도를 높였다. 이를 위해 6년간의 무인항공기 수리부속 소모 데이터를 수집하여 수요예측 분류모델을 제시하였다.

**Abstract** There has been a growing global interest in unmanned aerial vehicles (UAV) as strategic weapon systems in recent years. In line with this global trend, the South Korean military also focuses on maintaining combat readiness by developing UAVs. To effectively operate UAVs, the supply and maintenance of spare parts are essential, and accurate prediction of the demand for these spare parts is crucial. Thus far, the Korean military has performed spare part demand forecasting based on several time-series techniques using equipment maintenance information system data. On the other hand, an improved demand forecasting model is required to enhance the accuracy of spare part forecasting. This study compared and analyzed oversampling and undersampling models to propose a model that can improve the accuracy of spare part demand forecasting. Preprocessing techniques were also applied to enhance data distribution consistency and increase accuracy. To this end, six years of UAV spare part data were collected to present a demand forecasting classification model.

**Keywords** : UAV, Spare Parts, Demand Forecast, Data Mining, Deep Learning

### 1. 서론

무인항공기(UAV: Unmanned Aerial Vehicle 이하 UAV)라고도 알려진 드론은 조종사가 탑승하지 않고 원격제어 또는 사전 프로그래밍된 경로를 통해 자동 또는 반자동 비행이 가능하다. 또한 이러한 UAV는 인공지능을 통합하여 자체 환경 평가를 기반으로 자율 비행 결정을

내릴 수 있다.

이러한 특징들로 인해, 현대 사회에서 드론의 중요성이 증가하고 있다[1]. 드론은 군사, 법 집행, 수색 구조, 기계 관리, 자원 관리 등 다양한 분야에서 활용되고 있다 [2]. 군사작전에서는 적의 전술적 움직임과 전투력을 감시하고, 적의 전투력을 무력화하기 위해 사용된다[3]. 법 집행 부문에서는 사고 위치를 모니터링하고, 정보를 수

\*Corresponding Author : Jae-Dong Kim(Korea Institute For Defense Analyses)

email: soobahkin@gmail.com

Received May 9, 2023

Accepted June 2, 2023

Revised June 1, 2023

Published June 30, 2023

집하며 관련 정보를 전파하는데 사용될 수 있다[4].

수색 및 구조 분야에서는 사고 현장을 모니터링하고 데이터를 수집하며 잠재적 정보를 전달할 수 있다. 기계 관리 분야에서는 자원의 상태를 관찰할 수 있을 뿐만 아니라 관련 데이터를 수집하고 전송할 수 있다[5-7]. 현재 한국군은 시계열 방법을 중심으로 산술 평균법과 이동 평균법 등 5~8개의 방법을 활용하여 수리부속의 수요 예측에 대한 정확도를 높이기 위해 노력하고 있다[8].

효율적인 수요예측을 위해 빅데이터와 데이터마이닝 방법을 적용하여 민간 부문에서 활발히 연구되고 있다. 유사한 연구 결과들이 수리부속 데이터를 관리하는 한국군의 장비 유지관리 정보 시스템의 개발을 알려주고 있다[9].

민간 부문에서 관리되고 있는 수리부속 데이터와 관리 방법을 벤치마킹하여 전투력 향상과 예산 최적화라는 두 가지 중요한 목표를 달성할 수 있다[10].

본 연구의 목적은 UAV 수리부속의 데이터 균형을 통한 전처리기술을 사용하여 수리부속의 수요예측 정확도를 향상하는 것이다.

이 논문은 다섯 개의 장으로 구성하였다. 1장에서는 연구 배경과 목표를 제시하였으며 2장에서는 현재 연구와 관련된 선행연구를 분석한다. 3장에서는 연구데이터와 방법론의 개요를 제시하고 4장에서는 분석결과를 제시하였다. 마지막으로, 5장에서는 결론 및 향후 연구 방향을 제시한다.

## 2. 문헌연구

### 2.1 시계열 방법

시계열 데이터는 시간의 흐름에 따라 식별된 데이터로 과거에 관측된 데이터를 바탕으로 미래의 데이터 예측이 가능하다. 시계열 방법은 산술평균법(AMM: Arithmetic Mean Method 이하 AMM), 단순이동평균법(SMA: Simple Moving Average 이하 SMA), 가중이동평균법(WMA: Weighted Moving Average 이하 WMA), 선형이동평균법(LMA: Linear Moving Average 이하 LMA), 최소자승법(LSM: Least Square Method 이하 LSM) 등의 5가지 방법을 활용한다[11]. AMM은 수요를 예측할 때 과거 자료를 이용하여 발생한 수요를 모두 평균하여 예측하는 것으로 과거 자료가 충분히 많고, 균등하게 형성될 경우에 활용되는 방법이다. LSM은 전체 자료의 추세를 이해하는데 가장 많이 사용되며 관측치와

오차를 최소화하여 예측하는 방법이다[12]. 시계열 예측이 활용되는 사례는 아래와 같다. 기존 연구에서는 한국의 국내총생산 시계열 자료를 분석하여 2년간의 미래 예측에서 비관측요인모형이 보다 우수함을 확인하였고[13]. 시계열 데이터를 활용하여 코로나19 동향을 예측하여 미래의 질병에 대해서 방역과 방역 수칙의 중요함을 강조하였다[14]. 또한 여러 가지 시계열 모형을 활용하여 최대 전력수요를 예측하고 정확도를 높이는 방법을 연구하는 등 공공 분야에서도 선행연구가 이루어지고 있다[15].

### 2.2 데이터마이닝 방법

데이터마이닝은 방대한 양의 데이터를 탐색하여 의미 있는 패턴이나 규칙을 발견하기 위한 과정이며 본 연구는 다음과 같은 여러 데이터마이닝 방법을 사용한다.

의사결정나무(DT: Decision Tree 이하 DT)는 의사결정 규칙을 나무구조로 도표화하여 분류와 예측을 하는 데이터마이닝 방법이다. 기존 연구에서는 기업들의 효율적인 열처리 선택을 위해 DT 모델을 골격으로 델파이 방법과 AHP(Analytic Hierarchy Process) 방법을 사용하여 열처리선택 의사결정 모델을 개발 하였다.

랜덤 포레스트(RF: Random forest 이하 RF)는 과적합(overfitting) 문제와 같은 기존의 나무 모델(Tree Model)의 한계점을 개선하기 위한 앙상블 방법으로, 기존의 훈련 데이터로부터 추출한 부트스트랩 샘플(Bootstrap Samples)로 다양한 나무 모델을 형성하는 방법이다. 즉, 여러 나무 모델을 형성하는 과정에서 임의성(Randomness)을 부여하여 서로 상관되지 않은(Uncorrelated) 나무들을 형성하고 이들의 분류 결과나 회귀 결과를 종합하여 예측력을 향상하는 개념이다[16,17]. 기존 연구에서는 도시유역에 대한 실시간 홍수 지도를 예측하기 위해 기상청에서 제공한 강수량과 도시유역에 대한 2차원 침수해석 결과 그리고 기계 학습 중 하나인 회귀모델을 활용하여 실시간으로 도시유역에 대한 침수지도를 예측할 수 있는 방법을 제시하였다[18].

K-근접 이웃(KNN: K-nearest Neighbor 이하 KNN)은 근거리 데이터를 기반으로 데이터를 분류한다. 여기서 분류할 데이터에 가장 가까운 K 데이터가 포함된다. 이 방법은 데이터에 대한 분포 가정을 필요로 하지 않는 비모수적 방법에서 간단하고 직관적이며 정확도가 높기 때문에 널리 사용된다. 그러나 K 값에 대한 의존도는 높고 최적의 K 값을 찾는 것은 어려울 수 있다. 또한 표본 수가 많으면 계산 부하가 증가하고 모형 적합 시간

이 길어진다[19]. 기존 연구에서는 높은 엔트로피를 가진 그래프 기반 KNN 접근 방식을 사용하여 합금의 안정성을 예측했다. 그래프 기반 접근법을 사용하여 각 노드와 가장 가까운 노드 사이의 거리를 계산하여 엔트로피가 높은 합금의 안정성을 예측할 수 있다[20].

나이브 베이즈(NB: Naive Bayes 이하 NB)는 각 변수가 조건부 독립적이라고 가정할 때 가장 단순한 유형의 베이즈안 네트워크이다. 이 방법은 각 변수의 개별 조건부 확률을 고려하여 사건의 조건부 확률을 추정한다. NB는 다양한 분야에서 사용되고 있으며, 간단한 구조로 복잡한 문제를 해결할 수 있다[21].

서포트 벡터 머신(SVM: Support Vector Machine 이하 SVM)은 데이터를 고차원 공간에 매핑하여 최적의 의사 결정 경계를 결정한다. SVM은 다양한 커널 기능을 활용하여 다중 클래스 분류 및 회귀 분석문제를 해결할 수 있으며, 고차원 데이터에서 우수한 성능을 발휘한다. 차원 축소를 통해 더 나은 성능을 달성할 수 있으므로 분류 및 회귀 분석문제에 유용하다[22].

교각의 안전진단 예측을 위한 5가지 데이터마이닝 알고리즘의 성능을 비교한 최근 연구에서 서포트 벡터 머신이 가장 적합한 알고리즘으로 밝혀졌다[23].

선형회귀(LR: Linear Regression 이하 LR)은 관련 변수가 제공될 때 변수 간의 관계를 분석하고 결과를 예측하는데 사용되는 통계적 방법이다.

독립변수는 결과에 영향을 미치는 변수이고, 종속변수는 독립변수의 영향으로 발생하는 결과이다. 선형 회귀 분석은 독립변수와 종속변수 간의 관계에 따라 선형 또는 비선형으로 분류할 수 있다. 선형 회귀 분석에서 종속 변수는 독립변수가 증가하면 계속 증가하거나 감소한다[24].

### 2.3 딥러닝 방법

딥러닝은 모델 자체 내에 특정 추출 과정을 포함시켜 더 높은 정확도를 달성할 수 있는 방법이다. 딥러닝은 학습 과정에서 감지된 오류를 기반으로 매개 변수 값을 조정함으로써 예측력을 향상시킬 수 있다. 본 연구는 다음과 같은 여러 딥러닝 방법을 사용한다.

다층 퍼셉트론(MLP: Multilayer Perceptron 이하 MLP)은 입력 레이어, 숨겨진 레이어 및 출력 레이어가 있는 전방향(Feed-Forward) 신경망이다. 다양한 신경망 모델이 존재하지만, 다층 퍼셉트론 신경망은 데이터 분석에 가장 일반적으로 사용된다[25]. 기존 연구에서는 개별 스펙트럼 분석과 비선형 다층 퍼셉트론 네트워크를

결합한 하이브리드 지능형 알고리즘을 최적화하여 단기 부하 예측 모델을 제안한다. 이 방법론은 다양한 데이터 셋에서 유망한 성능을 보여주었다[25].

순환신경망(RNN: Recurrent Neural Network 이하 RNN)은 현재 입력 및 과거 상태 정보를 통합하여 미래 데이터를 예측하는 알고리즘이다. 전방향(Feed-Forward) 인공 신경망과 달리, RNN은 업데이트된 상태 정보가 다음 시간 차이의 뉴런을 통해 순환하는 반복 신경망의 일종이다. 이러한 특성 때문에 RNN은 일반적으로 시간 경과에 따른 순차적 특성을 갖는 데이터 분석에 사용된다[26]. 본 연구의 초점은 순환 신경망을 사용하여 기계 번역의 정확도를 향상시키는 것이다[27].

게이트 순환 장치(GRU: Gated Recurrent Unit 이하 GRU)는 LSTM(Long Short-Term Memory) 모델의 단순화된 변형이다. GRU에는 입력과 상태 간의 관계를 정의하는 업데이트 게이트와 상태 간의 관계를 정의하는 재설정 게이트가 포함된다. 업데이트 게이트는 이전 상태와 현재 입력 값을 기준으로 현재 상태를 업데이트하고, 재설정 게이트는 이전 상태를 잊은 다음 현재 입력 값을 기준으로 새 상태를 생성한다. GRU는 LSTM보다 비교적 단순한 구조를 가지고 있어 더 빠른 학습과 더 나은 성능을 제공한다[28]. 기존 연구에서 비선형 회귀 과정의 남은 유효 수명을 예측하기 위해 GRU 기반 RNN이 제안되었다. 이 모델은 정규화 방법을 사용하여 신경망 구조를 단순화하고 과적합을 방지하기 위해 비용 기능을 최소화한다. GRU 기반 RNN은 다른 모델보다 더 나은 결과를 보였다[28].

장단기 메모리(LSTM: Long Short-Term Memory 이하 LSTM)은 RNN에서 사용되는 메모리 셀을 개선하여 기울기 소실 문제를 해결한다. 기존 논문에서 LSTM 기법과 같은 딥러닝 기법이 주식가격이나 전력 수요예측 분야 등에 다양하게 활용되고 있다. 이러한 추세를 고려하여 한국 공군의 'B' 항공기 수리부속 수요예측에 LSTM 기법을 최초로 적용하고 성과를 분석하여 한국 공군 수리부속 수요를 예측하였다[29].

### 2.4 오버샘플링과 언더샘플링

오버샘플링은 불균형 데이터 셋을 해결하는데 사용되는 데이터마이닝 방법으로, 한 클래스는 다른 클래스보다 예제 수가 현저히 적다. 오버샘플링의 목표는 소수 클래스의 예제 수를 늘려 클래스 분포의 균형을 맞추는 것이다. 오버샘플링에서 소수 클래스에 대한 새로운 예제는 기존 예제를 기반으로 무작위로 복제하거나 새로운

예제를 생성하여 합성된다. 이를 통해 발생 또는 미발생 클래스를 다수 클래스에 유사한 수준으로 증가시켜 데이터마이닝의 정확도가 향상된다. 오버샘플링에는 무작위 오버샘플링(Random OverSampling), SMOTE 등 여러 가지 방법이 있다. 전반적으로 오버샘플링은 불균형 데이터 셋의 처리에 효과적이고 특히, 발생 또는 미발생 클래스가 충분히 표현되지 않은 경우에 더욱 효과적이다. 데이터마이닝 알고리즘이 의사 결정 경계를 보다 정확하게 학습할 수 있도록하고 다수 클래스에 대한 편향을 방지한다. SMOTE 알고리즘은 다음과 같다.

- SMOTE는 기존 객체와 유사한 새 객체를 만들어 소수 클래스에 대한 합성 샘플을 생성한다. 이 알고리즘은 소수 클래스에서 하나의 객체를 선택하고 가장 가까운 이웃을 찾아 낸다. 선택한 각 인접 라우터에 대해 원래 객체와 인접 라우터의 특성 값 간의 차이를 취하여 이 차이에 0과 1 사이의 난수를 곱하여 새 합성 객체를 만든다. 생성된 합성 객체가 소수 클래스에 추가된다. 이 과정은 소수 클래스가 다수 클래스와 균형을 이루거나 지정된 목표 균형 비율에 도달할 때까지 반복된다. SMOTE알고리즘은 아래와 같다.(Synthetic Minor Over-Sampling Technology)

Table 1. Pseudocode of Synthetic Minority Over-sampling Technique algorithm (SMOTE)

---

Input: original dataset, minority\_class\_label, k\_neighbors, oversampling\_ratio  
 Output: new\_dataset

Step 1. Retrieve all samples from the minority class in the dataset and store them in minority\_class\_data.  
 Step 2. Calculate the number of synthetic samples to generate based on the oversampling ratio and store the result in n\_synthetic\_samples.  
 Step 3. Find the k-nearest neighbors for each minority class sample using the KNN algorithm.  
 Step 4. Initialize an empty list called synthetic\_samples to store the newly generated synthetic samples.  
 Step 5. For i in range(n\_synthetic\_samples) do:  
 (a) Randomly select a minority class sample from minority\_class\_data.  
 (b) Find the k-nearest neighbors for the selected sample.  
 (c) Randomly select one of the k-nearest neighbors and store it in nn.  
 (d) Generate a new synthetic sample by interpolating between the selected sample and nn using the generate\_synthetic\_sample function.  
 (e) Add the synthetic sample to the list synth

---

언더샘플링은 데이터 셋의 클래스 불균형을 해결하기 위해 사용되는 또 다른 방법이다. 클래스 분포의 균형을 맞추기 위해 다수 클래스에서 일부 샘플을 제거하여 소수 클래스를 더 잘 학습할 수 있도록 한다. 그러나 이 방

법은 대다수 클래스 샘플의 일부 제거로 인해 정보가 손실될 수 있으며, 이는 분류기의 성능에 부정적인 영향을 미칠 수 있다. 따라서 언더샘플링은 학습에 사용할 수 있는 소수 클래스 샘플이 충분할 때 가장 효과적이다. 언더샘플링에는 랜덤 언더샘플링과 TomekLinks가 가장 일반적인 방법이다. 랜덤 언더샘플링은 다수 클래스 샘플의 무작위 제거를 수반하는 반면, TomekLinks는 다수 클래스와 소수 클래스의 경계에 가장 가까운 샘플을 식별하고 제거하는 방법이다. TomekLinks 알고리즘은 다음과 같다.

- TomekLinks는 서로 가깝지만 서로 다른 클래스에 속하는 객체 쌍을 식별한다. 구체적으로 그것은 그들 사이의 거리가 최소이고 객체가 다른 클래스에 속하는 객체 쌍을 찾는다. 이러한 쌍을 TomekLinks라 한다. 이 알고리즘은 TomekLinks의 일부인 객체를 노이즈 또는 경계선 객체로 간주하여 제거한다. 이 과정은 두 클래스 간의 더 깨끗한 분리를 만들어 데이터마이닝 알고리즘이 의사 결정 경계를 더 쉽게 학습할 수 있도록 한다. 이 방법은 단순한 무작위 언더샘플링 보다 더 많은 정보를 보존하고 중요한 정보의 손실을 방지한다.

Table 2. Pseudocode of TomekLinks algorithm

---

Input: dataset X with n examples, where each example xi has d features, and corresponding labels yi  
 Output: new dataset X' with noisy/borderline examples removed

Step 1. Compute the KNN graph for the dataset X using Euclidean distance, where k is a hyperparameter.  
 Step 2. For each example xi in X, determine if it is a TomekLink.  
 (a) If yi = 1, check if there exists any example xj in X such that yj = 0, d(xi, xj) is the shortest distance between xi and xj among all pairs of examples with different labels, and there are no other examples that have a shorter distance to xi or xj than the distance between xi and xj.  
 (b) If yi = 0, check if there exists any example xj in X such that yj = 1, d(xi, xj) is the shortest distance between xi and xj among all pairs of examples with different labels, and there are no other examples that have a shorter distance to xi or xj than the distance between xi and xj.  
 Step 3. Remove the noisy/borderline examples from X to obtain X'. The remaining examples are considered to be well-separated and informative.

---

본 연구에서는 SMOTE 오버샘플링 방법과 TomekLinks 언더샘플링 방법을 활용하여 모델 2와 모델 3의 성능을 각각 제시하였다. 또한, 그림 2는 오버샘플링과 언더샘플링의 개념을 그림으로 표현한 것이다.

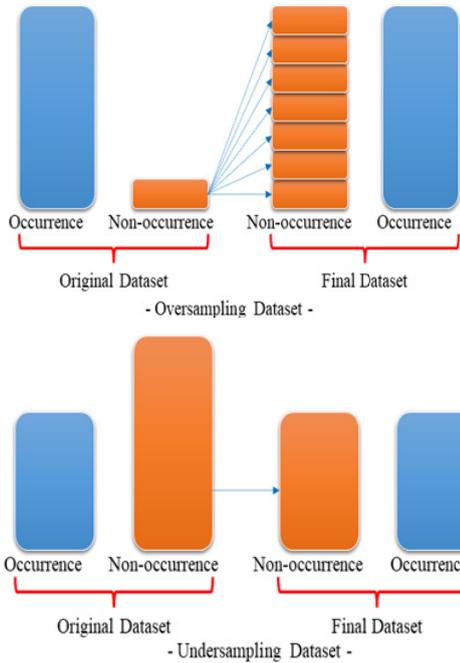


Fig. 1. A concept of oversampling and undersampling model

### 3. 수리부속 수요예측 모델 제한

#### 3.1 데이터 수집 및 변수 추출

군에서 활용되고 있는 장비정비정보체계(DELIIIS: Defense Logistics Integrated Information System 이하 DELIIS)는 2009년에 개발을 시작하여 2020년 재개발된 시스템으로 위의 [Table 3]과 같이 보급과 정비에 대한 자산을 관리하고 있다. 본 시스템은 편성 부대로부터 최

Table 3. Army DELIIS Table Example

Equipment Name	Equipment NSN	Equipment registration Number	maintenance date	NSN	troops
Equipment A	A0001	2001A1	20130514	A54321	X LSC
Equipment Z	C0001	2001C1	20210320	D54321	X battalion
...	...	...	...	...	...
Request Unit	consumption Unit	consumption	worker	number of workers	recoverability
EA	EA	20	최정호	3	N
EA	EA	10	홍길동	3	Y
...	...	...	...	...	...

상위 부대인 국방부에 이르기까지 통합 운용하고 있다. 본 연구에서는 육군의 전략자산 중 대대급 주력기인 UAV를 연구 대상 장비로 선정하였다.

육군 DELIIS의 정비테이블에서 53,422개 트랜잭션 데이터를 수집하였고, 수집된 데이터에는 정비날짜, 정비당일 소모된 수리부속품목의 개수, 대당 구성수, 표준단가, 조달기간, 비고(정비내역) 등 22개 항목이 포함되어 있다.

#### 3.2 변수 추출

본 연구의 목적은 수리부속 수요예측의 정확도를 향상시키는 것이다. 본 연구는 146개의 수리부속을 기반으로 변수를 추출하였다.

[Table 4]와 같이 2017년부터 2022년까지 6년간 11개의 구조화 데이터 변수를 연간 자료로 추출하였다. 대상 변수는 총 146개 품목 중 10,510건에 대한 수요가 관측된 품목별 2022년 수요 자료이다. 데이터 셋의 균형을 맞추기 위해 오버샘플링 방법중 SMOTE 알고리즘을 사용하여 수요가 있는 91개 품목과 수요가 없는 55개 품목을 과다 샘플링해 수요가 없는 품목을 36개에서 91개로 늘렸다. 그 결과, 추출된 데이터 셋은 182개의 항목으로 구성되었다. 또한 수요가 있는 수리부속의 수를 91개에서 36개로 줄이기 위해 언더샘플링 방법중 TomekLinks 알고리즘을 사용하여 총 110개의 항목이 생성되었다.

Table 4. Feature Description

	Feature	Means
independent variable	Maintenance Troop 1	Binary [0,1] data (Whether or not maintenance occurs in the maintenance unit)
	Maintenance Troop 2	
	Maintenance Troop 3	
	Maintenance Troop 4	
	Maintenance Troop 5	
	2017 year Consumption	Spare Parts By Item/Unit/Year Sum of consumption
2018 year Consumption		
2019 year Consumption		
2020 year Consumption		
2021 year Consumption		
dependent variable	Sum of Consumption	Consumption from 2022 year (Binary)

#### 3.3 모델 수립

본 연구의 목적은 기본, 오버샘플링, 언더샘플링의 결과를 비교 분석하여 데이터 균형이 모델 성능에 미치는 영향을 분석하는 것이다. 기본 모델은 데이터 균형조정

없이 개발되었으며, 오버샘플링 및 언더샘플링은 전처리 방법인 데이터 균형 조절을 사용하여 데이터 분포의 일관성을 향상시켜 데이터 분석결과의 정확성을 높였다.

이 과정은 그림 1과 같이 변수추출(Feature Selection), 데이터전처리(Preprocessing), 방법론(Learning Strategy), 평가척도(Performance Measure)의 4단계로 구성되어 있다. 첫째, 변수추출은 DELIIS에서 6년간의 UAV 데이터를 추출했다. 둘째, 데이터전처리는 기본, 오버샘플링 및 언더샘플링의 3가지 방법을 적용했다. 그중 오버샘플링은 SMOTE를 적용하였으며 언더샘플링은 TomekLinks를 적용하였다. 셋째, 방법론은 시계열 5가지, 데이터마이닝 6가지, 딥러닝 4가지 방법을 적용하였다. 마지막으로, 평가척도는 Accuracy, Recall, Precision, F1-Score를 기준으로 각 모델을 평가했다. 이 과정을 통해 UAV 수리부속에 대한 수요예측의 정확도가 향상되었다.

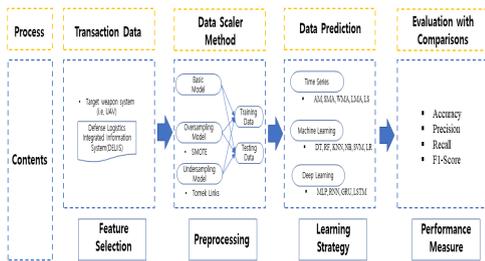


Fig. 2. Overview of the demand prediction process based on preprocessing.

본 논문의 목적은 데이터 균형 조절 없이 훈련된 모델 1과 데이터 균형 조절을 적용한 모델 2, 모델 3 세 가지 모델의 특징을 비교하고 데이터 균형 조절이 모델 성능에 미치는 영향을 검토하여 최적의 모델 선택 방법을 제시한다.

연구에 사용된 시계열 방법은 AMM, SMA, WMA, LMA, LSM이고 데이터마이닝 방법은 DT, RF, KNN, NB, SVM, LR이다. 또한, 본 연구는 MLP, RNN, GRU, LSTM을 포함한 딥러닝 방법을 사용한다. 분석프로그램은 Python 3.7.4를 사용하고 결과의 신뢰성과 정확성을 보장하기 위해 데이터 셋을 무작위로 추출하여 모델을 구성하고 테스트했다.

## 4. 실험 분석 및 결과

### 4.1 모델 평가

본 연구에서는 데이터마이닝을 적용한 모델의 효과를 평가하기 위해 오차행렬(Confusion Matrix)을 활용하였다. 이 평가 방법을 적용하여 Accuracy, Recall, Precision 및 F1-Score의 정확도를 측정하였다. 또한 모델의 성능을 평가하기 위해 훈련데이터 셋(Training data set)과 테스트 셋(Test data set)은 7:3 비율로 나누어 10겹 교차 검증(10-fold cross validation) 방법을 사용했다.

오차행렬의 예는 [Table 5]에 나와 있다.

Table 5. Confusion matrix

		Predicted	
		Y (Positive)	N (Negative)
Actual	Y (Positive)	A (True Positive)	B (False Negative)
	N (Negative)	C (False Positive)	D (True Negative)
Accuracy = $\frac{A + D}{A + B + C + D}$		Recall = $\frac{A}{A + B}$	
Precision = $\frac{A}{A + C}$		F1-Score = $\frac{2(Precision \cdot Recall)}{(Precision + Recall)}$	

### 4.2 모델 결과

모델 1, 모델 2, 모델 3의 Accuracy를 평가한 결과 아래의 [Table 6], [Table 7], [Table 8]과 같다.

모델 1은 시계열, 데이터마이닝, 딥러닝 방법의 결과를 제시하였다. 시계열 방법에서는 AMM과 SMA이 72%의 높은 Accuracy를 보여주었고 머신러닝 방법에서는 DT가 72%로 높은 Accuracy를 보였고 딥러닝에서는 RNN이 82%의 가장높은 Accuracy를 보였다. 딥러닝이 시계열과 데이터마이닝 보다 전체적으로 높은 Accuracy를 보였다. 그러나 특정 결과는 시계열과 데이터마이닝도 높은 결과를 보였다. LSM의 Accuracy는 62%로 가장 낮았지만 F1-Score가 77%로 가장 높아 Accuracy와 Recall간의 균형이 잘 잡혀있음을 알 수 있었다. SVM과 LR의 Accuracy는 62%로 가장 낮았지만 Recall이 100%로 관련된 모든 객체를 탐지할 수 있지만 잘못된 긍정의 수가 많다는 것을 알 수 있었다. 전반적으로 본 연구는 RNN이 Accuracy, Precision, Recall, F1-Score 측면에서 다른 모델보다 더 나은 성능을 보였다.

Table 6. Model 1 result of time series, data mining and deep learning

		Accuracy	Recall	Precision	F1-Score
Time Series	AMM	72%	85%	72%	75%
	SMA	72%	85%	72%	75%
	WMA	68%	89%	68%	75%
	LMA	71%	88%	74%	74%
	LSM	62%	89%	62%	77%
Data Mining	DT	72%	80%	72%	74%
	RF	69%	85%	69%	74%
	KNN	70%	85%	70%	74%
	NB	70%	85%	70%	74%
	SVM	62%	100%	62%	77%
	LR	62%	100%	62%	77%
Deep Learning	MLP	68%	80%	68%	71%
	RNN	82%	82%	82%	82%
	GRU	79%	81%	79%	80%
	LSTM	79%	81%	79%	80%

모델 2는 오버샘플링을 적용한 데이터마이닝과 딥러닝 방법의 결과를 제시하였다. 데이터마이닝 방법에서는 DT와 SVM이 75%의 Accuracy를 보였고 딥러닝에서는 GRU와 LSTM이 84%의 높은 Accuracy를 보였다. 딥러닝에서 LR은 100%의 Recall 점수를 보여 모든 실제 긍정적인 객체가 정확하게 식별되었음을 나타냈다. 마지막으로 LSTM은 F1-Score가 84%로 가장 높았으며 이는 Precision과 Recall에서 균형잡힌 성능을 나타냈다. 딥러닝이 데이터마이닝 보다 전체적으로 우수한 결과를 보였다. 모델 1과 비교하면 데이터마이닝은 유사한 결과가 도출되었으나 딥러닝의 경우 RNN이 가장 낮은 값을 보였고 MLP, GRU, LSTM의 결과 값은 상승하였다.

Table 7. Model 2 Machine learning result using oversampling technique

		Accuracy	Recall	Precision	F1-Score
Data Mining	DT	75%	80%	75%	76%
	RF	72%	80%	72%	74%
	KNN	71%	82%	71%	74%
	NB	70%	85%	70%	74%
	SVM	75%	82%	75%	77%
	LR	62%	100%	62%	77%
Deep Learning	MLP	80%	81%	80%	80%
	RNN	78%	78%	78%	78%
	GRU	84%	84%	84%	84%
	LSTM	84%	85%	84%	84%

모델 3은 언더샘플링을 적용한 데이터마이닝과 딥러닝 방법의 결과를 제시하였다. 딥러닝의 Accuracy는 RNN을 제외하고 데이터마이닝 보다 높은 값을 보였으며 그중 MLP가 77%로 가장 높았다. 전반적으로 MLP가 가장 높은 값을 보인 반면 RNN은 가장 낮은 값을 보였다. 또한 LSTM을 제외한 모든 Recall이 Accuracy 보다 일관되게 높았으며 이는 모델이 부정적인 사례를 정확하게 제외하는 것보다 긍정적인 사례를 더 정확하게 식별한다. Accuracy와 Recall는 F1-Score 보다 일관되게 높았으며 이는 모델의 측정값의 균형을 맞출수 있음을 의미한다. 전반적으로 데이터마이닝에서는 유사한 값을 보였지만 딥러닝에서는 모델의 성능에 더 큰 변동이 있음을 시사했다. 전반적으로, 오버샘플링 방법을 사용한 기계 학습은 다른 모델보다 더 나은 결과를 보여주었습니다. 그러나 적절한 샘플링 방법을 선택할 때는 모델 성능에 상당한 영향을 미칠 수 있으므로 데이터의 특성을 고려하는 것이 중요하다.

Table 8. Model 2 Machine learning result using undersampling technique

		Accuracy	Recall	Precision	F1-Score
Data Mining	DT	71%	77%	63%	62%
	RF	72%	77%	64%	63%
	KNN	72%	77%	64%	63%
	NB	71%	77%	63%	62%
	SVM	72%	77%	64%	63%
	LR	72%	77%	64%	63%
Deep Learning	MLP	77%	78%	72%	73%
	RNN	71%	72%	63%	63%
	GRU	75%	75%	71%	72%
	LSTM	75%	74%	72%	73%

Time Series과 Data Mining을 비교했을 때, Time Series는 하나의 데이터 패턴을 적용하지만 Data Mining은 다양한 데이터 패턴을 적용할 수 있으므로 정확도가 높습니다. Data Mining과 Deep Learning을 비교했을 때, 딥러닝은 더 많은 변수를 생성하여 모델에 적용하기 때문에 정확도가 높습니다. 따라서 Time Series < Data Mining < Deep Learning 순으로 갈수록 데이터 패턴과 많은 변수를 고려하기 때문에 대체적으로 높은 정확도를 보입니다.

예를 들어, Time Series는 주식 시장의 가격을 예측하는 데 사용될 수 있습니다. Time Series는 과거의 가격 데이터를 사용하여 미래의 가격을 예측합니다. 그러

나 Time Series는 미래의 가격을 정확하게 예측할 수 없습니다. 왜냐하면 미래의 가격은 과거의 가격 데이터에 의해 결정되지 않기 때문입니다.

Data Mining은 Time Series보다 더 많은 데이터 패턴을 적용할 수 있습니다. Data Mining은 과거의 가격 데이터뿐만 아니라 기타 데이터, 예를 들어 경제 지표, 날씨 데이터, 뉴스 기사 등을 사용하여 미래의 가격을 예측합니다. 따라서 Data Mining은 Time Series 보다 더 정확한 예측을 할 수 있습니다.

## 5. 결론

본 논문은 육군이 운영하는 대대 UAV의 수리부속 수요 예측 제고를 위하여 지난 6개년의 수리부속 수요 데이터를 분석하고 전처리 기술로 데이터 분석결과와 정확도를 높였다. 전처리는 기본모델, 오버샘플링, 언더샘플링 방법을 사용하였다.

그 결과 오버샘플링을 적용한 딥러닝 방법을 활용한 수리부속 수요예측 모델을 제안한다.

오버샘플링을 적용한 딥러닝 방법중 GRU와 LSTM은 84%로 가장 높은 Accuracy 값을 보였다. 제안한 모델은 기본 모델에 비해서 5% 향상된 정확도를 나타냈다는 점에서 논문의 의미가 있다. 하지만 본 방법론은 수요예측 발생 여부인 분류문제도 정확도를 제시하였으므로 한계가 존재한다. 따라서 수량 예측이 가능한 모델로 추후 연구가 필요하다. 또한 장비 유지관리 특성, 거리, 사용연수 등 정형 및 비정형 데이터를 활용하면 수요예측의 정확도가 더욱 향상될 것으로 예상된다.

## References

- [1] Chaurasia, R., Mohindru, V. (2021). Unmanned aerial vehicle (UAV): A comprehensive survey. *Unmanned Aerial Vehicles for Internet of Things (IoT) Concepts, Techniques, and Applications*, 1-27.
- [2] Ahmed, F., Mohanta, J. C., Keshari, A., Prakash, P. (2022). Recent Advances in Unmanned Aerial Vehicles: A Review. *Arab Journal of Science and Engineering*, 47(12), 7963-7984.
- [3] Glade, D. (2000). *Unmanned Aerial Vehicles: Implications for Military Operations*. Center for Strategy and Technology, Air War College, Air University.
- [4] Dafinova, M., Martin, A. K., Taylor, L. (2019). *Police Surveillance Drones and Their Impact on the Right to Privacy*. *Journal of Information Policy*, 9, 347-374.
- [5] Mendis, N. D. N. A., Dharmarathne, T. S. S., Wanasinhe, N. C. (2017). Use of Unmanned Aerial Vehicles in Crime Scene Investigations-Novel Concept of Crime Scene Investigations. *Forensic Research Criminology International Journal*, 4(1), 1-2.
- [6] Jeong, K., Park, H., Lee, S. (2022). A Synthetic Review of UAS-Based Facility Condition Monitoring. *Drones*, 6(1), 420.
- [7] Mishra, P. K., Rai, A. (2021). Role of Unmanned Aerial Systems for Natural Resource Management. *Journal of the Indian Society of Remote Sensing*, 49, 671-679.
- [8] Gülan, M. A., Atay, A. T. (2010). Forecasting Spare Parts Demand Using Time Series Analysis and Autoregressive Integrated Moving Average Models. *International Journal of Production Research*, 48(9), 2517-2532.
- [9] Al-Jabri, S., Al-Jabri, M., Al-Jabri, M. (2017). Big Data and Data Mining Techniques for Effective Demand Forecasting. *International Journal of Computer Applications*, 168(5), 23-29.
- [10] Treiblmaier, A. (2022). Improving Efficiency Through Data-Driven Decision-Making in a Military Environment. *Defence Horizon Journal*, 11, 11-22.
- [11] Franses, P.H.; Legerstee, R. *Time Series Models for Business and Economic Forecasting*. Cambridge University Press: Cambridge, UK, 1998.
- [12] Lim, B., Zohren, S. (2021). Time-Series Forecasting with Deep Learning: A Survey. *Philosophical Transactions of the Royal Society A*, 379, 20200209.
- [13] Petropoulos, F., Makridakis, S. (2022). *Forecasting: Theory and Practice*. *International Journal of Forecasting*, 38(4), 1409-1411.
- [14] Alloghani, M., Al-Jumeily, D., Hussain, A. J., Fergus, P. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. In *Supervised and Unsupervised Learning for Data Science* (pp. 3-21). Springer.
- [15] Ierace, S., Cavalieri, S. (2013). An analytic hierarchy process based model for the selection of decision categories in maintenance systems. *Management and Production Engineering Review*, 4(2), 37-49.
- [16] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- [17] Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
- [18] Lee, S., Choi, S., Kim, M., Jang, J. (2017). Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomatics, Natural Hazards and Risk*, 8(4),

1185-1203.

- [19] Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.
- [20] Ghouchan Nezhad Noor Nia, R.; Jalali, M.; Houshmand, M. (2022). A Graph-Based k-Nearest Neighbor (KNN) Approach for Predicting Phases in High-Entropy Alloys. Applied Sciences, 12(16), 8021.
- [21] Campbell, C.; Ying, Y. (2011). Learning with Support Vector Machines. Morgan & Claypool Publishers, San Ra-fael, CA, USA.
- [22] Bao, Y.; Zhang, H.; Ou, J.; Chen, Y.; Yang, J.; Xiang, H.; Wang, Y. (2013). Damage Detection of Bridge Structure Based on SVM. Mathematical Problems in Engineering, 2013, 148070.
- [23] Chow, S.-C. (2018). Advanced Linear Models: Theory and Applications. Routledge, London, UK.
- [24] Niu, M.; Wang, Y.; Sun, C. (2016). An Innovative Integrated Model Using the Singular Spectrum Analysis and Nonlinear Multi-Layer Perceptron Network Optimized by Hybrid Intelligent Algorithm for Short-Term Load Fore-casting. Applied Mathematical Modelling, 40, 4079-4093.
- [25] Petneházi, G. (2019). Recurrent Neural Networks for Time Series Forecasting. arXiv preprint, arXiv:1901.00069.
- [26] Datta, D.; Bhattacharjee, D.; Ghosh, S. (2020). Neural Machine Translation Using Recurrent Neural Network. In-ternational Journal of Engineering and Advanced Technology, 9(5), 1395-1400.
- [27] Chen, J., Jing, H., Chang, Y., Liu, Q., (2019). Gated Recurrent Unit Based Recurrent Neural Network for Remaining Useful Life Prediction of Nonlinear Deterioration Process. Reliability Engineering & System Safety, 185, 372-382.
- [28] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long Short Term Memory Networks for Anomaly Detec-tion in Time Series. ESANN, 89-94.
- [29] Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. Information, 14(1), 54.

최 정 호(Jung-Ho Choi)

[정회원]



- 2004년 3월 : 육군3사관학교 환경 공학과 (공학사)
- 2013년 2월 : 경희대학교 경영학 과 (경영학석사)
- 2017년 8월 : 광운대학교 방위사 업학과 (박사수료)
- 2004년 3월 ~ 현재 : 육군 소령

<관심분야>

드론, 방위산업, 데이터마이닝

유 지 훈(Ji-Hoon Yu)

[정회원]



- 2006년 12월 : 미국 해군대학원 안전보장학과 (안전보장학석사)
- 2013년 8월 : 미국 시라큐스대 정 치학과 (정치학박사)
- 2021년 1월 ~ 2021년 12월 : 해군사관학교 군사전략학 교수
- 2022년 1월 ~ 현재 : 한국국방연 구원

<관심분야>

해양안보, 한미동맹, 외교/안보

김 재 동(Jae-Dong Kim)

[정회원]



- 2005년 2월 : 한양대학교 산업공 학과 (공학사)
- 2007년 2월 : 한양대학교 산업공 학과 (공학석사)
- 2020년 2월 : 고려대학교 산업경 영공학과 (박사수료)
- 2007년 5월 ~ 현재 : 한국국방연 구원 연구위원

<관심분야>

데이터마이닝, 최적화, 시뮬레이션, 국방정책