

국내 항공기 위치 데이터를 활용한 이착륙 접근 단계에서의 항공 위험상황 탐지를 위한 데이터 전처리 및 머신 러닝 분석 기법

¹ 이상호, ² 손일락, ³ 정규호, ^{4*} 박노삼

Data Preprocessing and ML Analysis Method for Abnormal Situation Detection during Approach using Domestic Aircraft Safety Data

¹Sang Ho Lee, ²Ilrak Son, ³Kyuho Jeong and ^{4*}Nohsam Park

요약

본 논문에서는 2019년 국내 공항을 기준으로 측정된 시계열 항공기 위치 데이터를 활용하여 국내 공항에 이착륙 시 접근 단계에서의 항공 위험상황 중 Go-Around 및 UOC_D 를 분석하고, 다양한 클러스터링 기반 머신 러닝 기법을 적용하여 국내 항공 데이터에서 가장 알맞은 분석 기법이 무엇인지를 실험을 통해 제시한다. 항공기 위치를 측정하기 위한 센서는 ADS-B를 단일로 사용하였으며, 클러스터링 기법들 중 K-Means, GMM, DBSCAN 등의 알고리즘을 사용하여 이상상황을 분류하기 위한 모델을 설계하였다. 그 중 해외에서는 RF 모델이 가장 나은 성능을 보였으나, 국내 항공 데이터에 대해서는 국내 지형에 특화된 부분을 반영한 GMM이 가장 높은 분류 성능을 보이는 것으로 실험을 통해 확인하였다.

Abstract

In this paper, we utilize time-series aircraft location data measured based on 2019 domestic airports to analyze Go-Around and UOC_D situations during the approach phase of domestic airports. Various clustering-based machine learning techniques are applied to determine the most appropriate analysis method for domestic aviation data through experimentation. The ADS-B sensor is solely employed to measure aircraft positions. We designed a model using clustering algorithms such as K-Means, GMM, and DBSCAN to classify abnormal situations. Among them, the RF model showed the best performance overseas, but through experiments, it was confirmed that the GMM showed the highest classification performance for domestic aviation data by reflecting the aspects specialized in domestic terrain.

Keywords: Machine Learning, Clustering, Gaussian Mixture Model, ADS-B, Anomalies, Go-around

¹ 주식회사 애나, CTO (shlee@byanna.io)

² 주식회사 애나, AI 연구센터, 선임연구원(ilrak.son@byanna.io)

³ 주식회사 애나, AI 연구센터, 연구원 (kyuho.jeong@byanna.io)

^{4*} 한국전자통신연구원 지능형휴먼트윈연구센터, 책임연구원, 교신저자(siru23@etri.re.kr)

I. 서론

우리나라의 항공교통은 여객과 화물 이외에도 다양한 목적의 항공 수요의 증가가 지속해서 이어지고 있어 수많은 항공기와 관련 항공산업의 성장이 이어지고 있다. 이러한 항공산업의 호황에 따라 사고 및 안전 장애 발생량도 증가하고 있어 항공안전 사고 방지를 위해 예방적인 기술이 필요하게 되었다. 기존 연구[1]에 따르면 미국, 유럽 등 항공 선진국들은 수십 년 전부터 항공 데이터 분석을 통한 증거기반 항공안전관리 정책을 수용 및 이행하고 있으며 미국의 경우 연방항공청(FAA)과 항공우주국(NASA)은 업무 협약을 통해 항공안전보고시스템(Aviation Safety Reporting System, ASRS)를 운영하고 있으며 수집된 항공 데이터를 기반으로 항공 안전사고 방지를 위해 이상 징후를 탐지하기 위해 알고리즘 기반의 이상 징후 예측 연구를 진행하고 있다.

우리나라에서도 안전사고 예방 및 방지를 위하여 국토교통부, 공항공사, 항공사 등 항공안전 관련 기관에서 기관별 관련 데이터를 수집 저장 관리하고 있다. 하지만, 다양한 이유로 해당 데이터들을 활용한 국내 환경에 특화된 이상 징후 분석 및 예측 알고리즘에 관한 연구는 큰 진전이 없는 것이 사실이다. 이에 본 논문에서는 국내 공항에 대해 접근 단계에 있는 이착륙 항공기들에 대한 ADS-B 센서 기반의 위치 데이터를 활용하여 이상 상황 중 Go-Around 및 UOC-D를 머신러닝의 클러스터링 알고리즘을 통해 분석하였으며, 해외 공항 및 국내 공항 데이터에 대해 각각 최적의 알고리즘이 무엇인지를 실험 결과를 통해 제시한다. 이를 위해 2 장에서는 사전 연구에 대해 소개하고, 3 장에서는 국내 공항에서 취득하는 ADS-B 데이터의 특성에 대해 설명하며 4 장에서는 국내 항공기의 ADS-B 데이터를 정제하는 방안에 대해 설명한다. 5 장에서는 국내 항공 이착륙 환경에서의 항공기 이상 징후를 정의하고, 6 장에서는 머신러닝 기반 항공 이상상황 분석 모델 선정 및 각 모델별 성능 실험 및 결과를 도출하며, 마지막으로 7 장에서 결론을 제시한다.

II. 사전 연구

국내 데이터를 활용하지 않고, 해외의 공개된 항공 데이터를 기반으로 하는 이상 징후 분석에 관하여 다양한 연구가 시도됐다. 기본적으로 데이터 수집부터 정제를 통한 머신러닝 모델 적용에까지 이르기 위한 프로세스는 Figure 1 에서 제시하는 바와 같으며, 데이터 수집 후 보간 과정을 통해 전처리(정제) 과정을 거쳐 원시데이터를 만들게 되며, 해당 데이터를 클러스터링 또는 특징점 추출을 통한 Regression 분석을 통해 이상상황에 대한 분석을 머신러닝 기법으로 실시하게 된다. 이후 각각의 머신러닝 분류 기법으로 도출된 결과값에 대해 정확도를 측정하는 방식으로 성능수치를 측정하여 기존연구와 비교하는 방법을 사용한다.

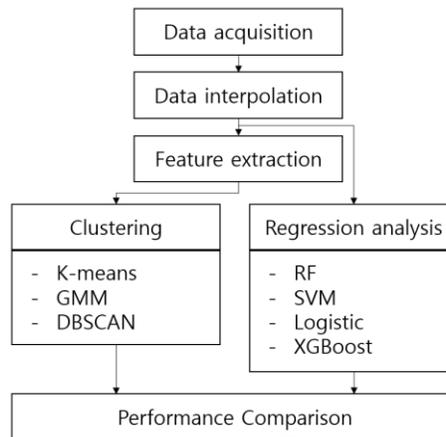


Figure 1. Comparison and Process for Aviation Anomaly Analysis Models

그림 1. 항공 이상징후 분석을 위한 모델 비교 및 프로세스

상기 프로세스에 기반한 기존 연구[2]에서는 항적 데이터를 활용하여 통계적 분석으로 이상 징후를 탐지하는 연구를 진행해 오고 있었으며, Dai 연구팀은 [3]에서 Go-around 탐지연구에서 항적 데이터로 머신러닝 알고리즘을 통한 이상 징후 분류 연구를 통해서 알고리즘별 이상 징후 탐지 정확도를 비교하는 결과를 제시하였다. 또한, Proud 연구팀은 [4]를 통해 TADSS 라는 비정상 궤적 탐지연구를 통해서 Figure 2 와 같이 이상 징후를 분류하는 결과를 도출한 바 있다. 하지만 이러한 연구들은 본 논문에서 제시하는 연구와는 원본 항적 데이터를 가공하지 않고 사용한다는 차이가 존재한다.

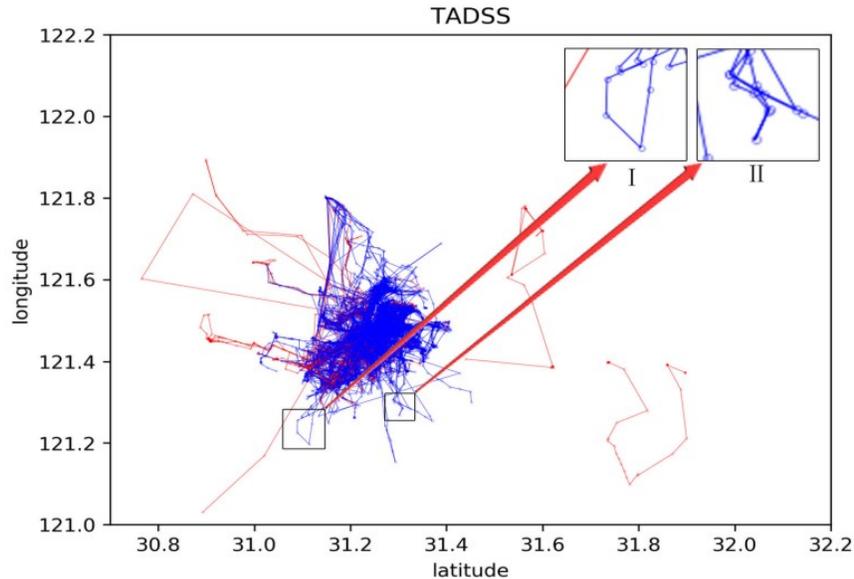


Figure 2. Anomaly trajectory detection result based on location data. Normal is blue, abnormal is red
그림 2. 위도, 경도 기반의 이상상황 항적 데이터. (정상 : 파란색, 비정상 : 빨간색)

III. 국내 항공 데이터 특성

국내 항공 데이터는 해외 항공 데이터와 비교하면 항공편 수가 현저히 적다. Figure 3 에서 보이는 바와 같이 최근 3 개년 공항별 항공편 수 그래프로 국내 주요 공항인 인천공항(RKSI)과 미국의 주요 공항인 존 F.케네디 공항(KJFK)의 2021 년도 항공편 수를 비교해 보면 연평균 150,000 개의 항공편 수가 차이가 나며 국가별로 항공편을 비교하게 된다면 그 차이는 더욱 벌어진다.

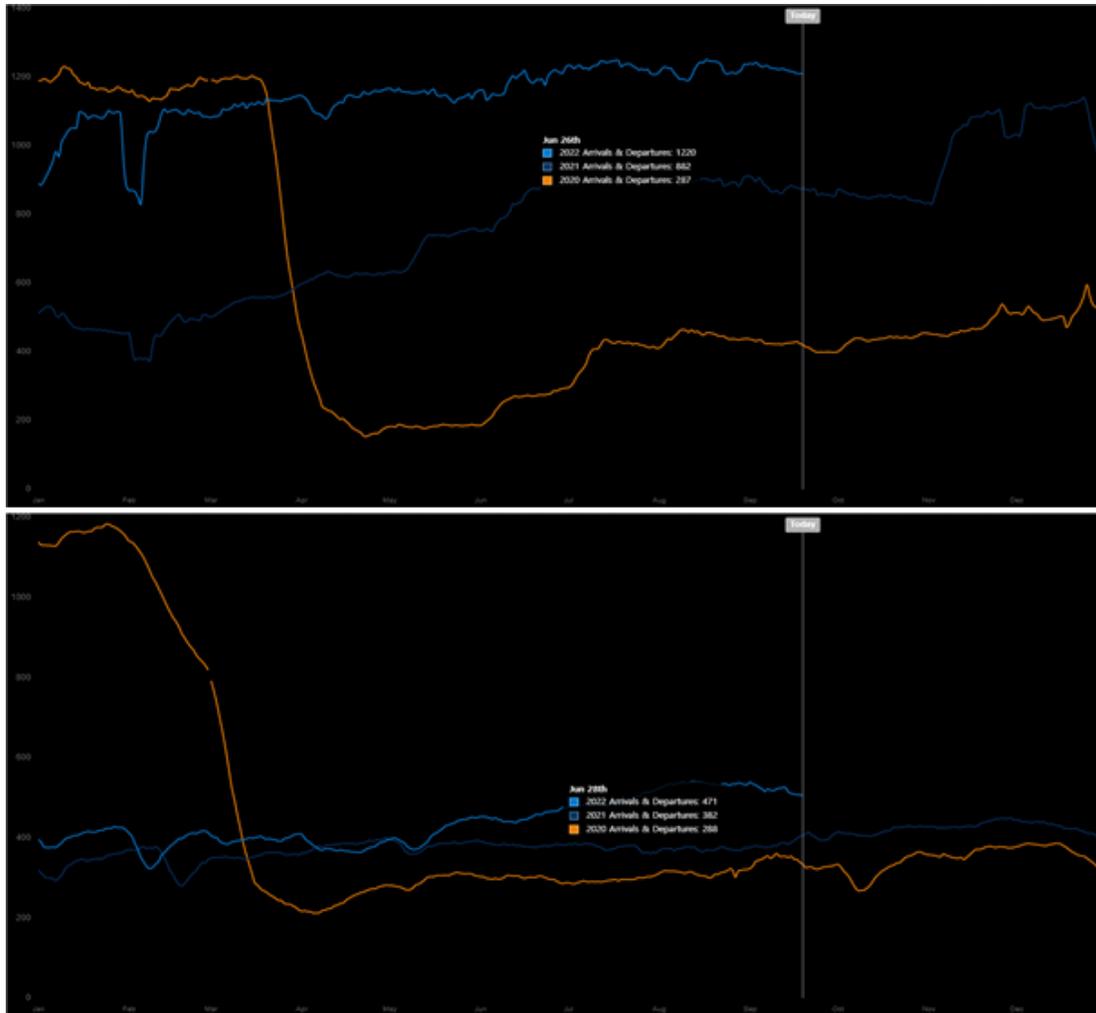


Figure 3. Major flight trends over 3 years (JFK, RKSI)

그림 3. 최근 3년간 미국 공항(JFK)과 국내 공항(RKSI)에 대한 항공편 수 차이 그래프

데이터 수의 차이는 이상 징후의 발생 데이터 수의 차이로도 직결된다. 항공편 10만 건 당 약 200 건의 이상 징후가 발생하는데 국내 항공 데이터에서 확인되는 이상 징후는 약 200 건으로 해외 항공 데이터에서의 이상 징후와 비교하면 오차가 발생할 가능성이 크다. Table 1.은 2020년도 주요 항공사 별 항공편 수이다.

Table 1. Number of flights by airline in 2020

표 1. 2020년에 대한 공항별 항공편 수

Airport name	flight
JFK	298,761
DFW	613,843
LAS	408,944
RKSI	148,621

또 해외 항공 데이터는 공항에 이착륙하는 항공기 데이터만 존재한다. 하지만 국내 항공 데이터에는 항공기 이외에 공항에서 사용한 모든 수직 이착륙기, 무인 드론의 항공 데이터도 포함되어 있다. 수직 이착륙기 및 무인 드론의 항공 데이터는 일반적인 항공기의 항적과 전혀

다른 양상을 보여 모든 데이터가 항공기의 이상 징후로 나타난다. 이러한 데이터는 이상 징후 분석 모델의 성능을 매우 악화시키는 요인이다.

Cho 연구팀의 기존 연구[5]에 따르면, 국내 공항은 바다와 인접한 공항이 대부분으로 윈드시어와 같은 기상 악화로 인한 이상 징후가 발생한다. 해외 항공 데이터는 이러한 기상 악화가 거의 발생하지 않기 때문에 해외 항공 데이터와 이상 징후의 특징이 다르게 나타날 수 있다.

항공 데이터는 radar 데이터와 ADS-B 데이터, RTK 데이터가 있는데 각각의 데이터는 수신 간격이 이론적으로 radar 데이터가 4 초, ADS-B 데이터가 0.5 초, RTK 데이터가 0.2 초[6]로 Figure 4 에서 보는 바와 같다. Figure 5 는 ADS-B 데이터의 수집 과정으로 해외 항공 데이터는 0.5 초 간격으로 수집되는 ADS-B 데이터를 사용하여 항공기의 항적을 추적하고 이상 징후를 탐지한다[7]. 국내 항공 데이터도 ADS-B 데이터를 사용하지만, 국내 ADS-B 데이터는 수신 간격이 해외 항공 데이터와 달리 평균적으로 20 여초 내외이며 이마저도 일정하지 않다. 따라서 국내 항공 데이터는 전처리 가공을 진행하지 않는다면 해외 데이터를 기반으로 사용한 이상 징후 분석 모델과 국내 데이터를 활용한 이상 징후 분석 모델을 비교하는 것은 무의미하다.

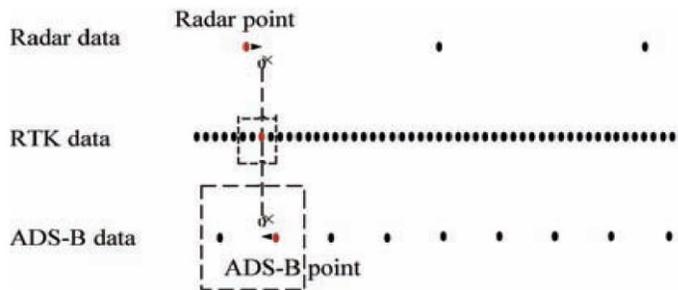


Figure 4. Time Interval and Synchronization Method According to Sensor Type
 그림 4. 센서별 데이터 취득 시간 간격 및 동기화 방법

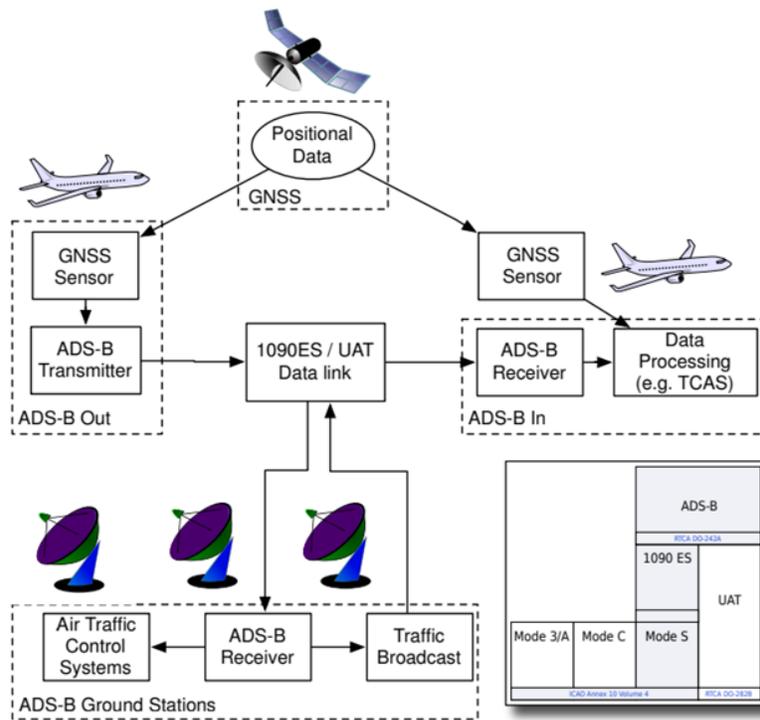


Figure 5. ADS-B system architecture and protocol hierarchy
 그림 5. ADS-B 시스템 아키텍처와 프로토콜 계층

IV. 국내 항공 이착륙 환경에서의 항공기 이상 징후

4.1 이상징후 정의

해외 항공연구기관에서는 항공기 이상 징후를 항적을 기반으로 가장 많이 나타나는 정상 경로를 제외하고 그 외로 많이 나타나는 항적들로 세분화하여 정의하고 이를 탐지하기 위해 사전 연구와 같은 이상 징후 탐지 및 분석 모델을 연구하고 있다[1]. 하지만 국내 항공 데이터는 해외 데이터와 상기와 같은 다른 특성이 있어 사전 연구와 다른 기준으로 연구를 진행해야 할 필요가 있다.

앞서 말한 바와 같이 국내 항공 데이터는 해외 항공 데이터와 달리 지형적인 특징으로 인한 이상 징후가 나타나며 이상 징후의 전체 발생 건수가 적어 이상 징후를 세분화하여 분류하기가 어렵다. 이에 따라 국내 항공 환경에서 발생하는 이상 징후를 Go-around(복행 및 실패접근) 및 UOC_D(예측불가 운항상황)으로 구분하고 다음과 같이 정의하고 있다.

- UOC_D: 항공기 접근구간에서 300ft 이상 상승한 경우
- Go-around: UOC_D 이상 징후로 탐지된 데이터 내에서 활주로 끝단으로부터 7NM 이내에서 2000ft 이상 상승한 경우.

4.2 국내 항공 데이터 특성을 고려한 이상징후 데이터 전처리

국내 항공 데이터에는 이전 데이터와의 수신 간격이 5분 이상 차이가 존재하였다. 이러한 데이터는 이상 징후 발생 여부를 확인하기가 어려워 ADS-B의 수신 시간 데이터를 확인하여 삭제하였다. 또한, 항공기와 다른 항적을 보이는 수직 이착륙기 및 무인 항공 데이터는 항공기 이상 징후로 분류되는 데이터로 분석 결과에 악영향을 미친다. 따라서 GPS 데이터가 이전 데이터와 비교하였을 때 위 경도의 차이가 거의 없으나 고도 차이가 높게 나타나는 데이터를 탐지하여 삭제하였다. Figure 6은 수신 간격이 5분 이상인 ADS-B 데이터이다. 이처럼 동일 항공편에서 수신 간격이 5분 이상 존재하는 데이터를 확인하고 삭제하였다. Figure 7은 헬기의 항적 데이터를 시각화한 그래프이다. 기존 항공기의 항적 데이터와 달리 수직 상승 곡선이 매우 가파른 것을 확인할 수 있다.

2:32:33 PM	ESR884	B738	33.88081	126.4755	8225
2:33:06 PM	ESR884	B738	33.91795	126.5044	8675
2:33:23 PM	ESR884	B738	33.93494	126.5183	9400
2:59:28 PM	ESR884	B738	36.64648	127.0749	9625
2:59:52 PM	ESR884	B738	36.66883	127.1072	9175
3:00:22 PM	ESR884	B738	36.6969	127.1456	8450

Figure 6. Domestic aircraft ADS-B data (time, callsign, type, latitude, longitude, altitude)
 그림 6. 국내 항공기의 ADS-B 시간 데이터(time, callsign, type, latitude, longitude, altitude)

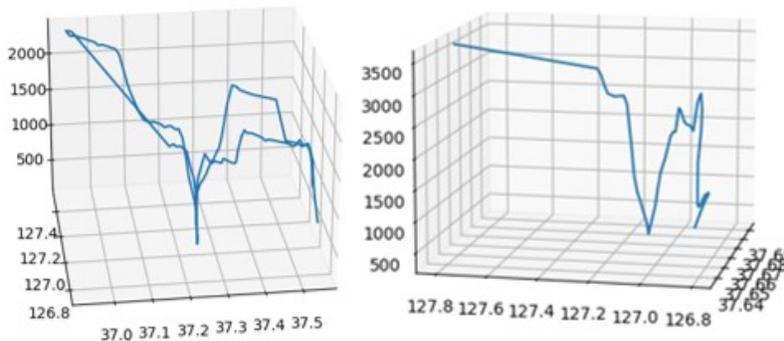


Figure 7. Vertical take-off and landing aircraft track data visualization graph
 그림 7. 수직 이착륙 항공기의 항적 데이터 시각화 그래프

국내 항공 ADS-B 데이터는 수신 간격이 0.5 초인 해외 데이터와 달리 평균 20 여 초의 간격으로 수집되며 전체적인 수신 간격이 일정하지 않다. 국내 항공 데이터를 해외 ADS-B 데이터를 기반으로 이상 징후를 분석한 모델과 성능 비교를 위해 유사한 데이터 형태를 만들어 주기 위해 국내 항공 ADS-B 데이터의 보간을 진행하였다. 또한, 기존의 시계열 항공 데이터는 UOC_D 이상 징후가 포함된 한국형 항공기 이상 징후 분석 모델에 사용하기에 적합하지 않아 항공 데이터를 기반으로 이상 징후를 탐지할 수 있도록 항공기 접근 단계에서의 고도 상승량, 고도 상승 시 수직 상승 최고 속도, 고도 상승 시 수직 상승 평균 속도의 특징 데이터를 추출하여 항공편 별 하나의 데이터로 치환하였다. Figure. 8 은 국내 항공 ADS-B 데이터를 보간한 결과이다.

Date	Time	Latitude_R	Longitude	Altitude_R	Date	Time	Latitude_P	Longitude	Altitude_P
2018-01-02	5:33:15	37.32527	126.581	2650	2018-01-02	5:33:15	37.32527	126.581	2650
2018-01-02	5:33:21	37.32816	126.5785	2600	2018-01-02	5:33:16	37.32575	126.5806	2642
2018-01-02	5:33:31	37.33511	126.5724	2550	2018-01-02	5:33:17	37.32623	126.5802	2633
2018-01-02	5:33:41	37.3412	126.5669	2500	2018-01-02	5:33:18	37.32672	126.5798	2625
2018-01-02	5:33:51	37.34728	126.5616	2350	2018-01-02	5:33:19	37.3272	126.5793	2617
2018-01-02	5:34:01	37.35305	126.5564	2200	2018-01-02	5:33:20	37.32768	126.5789	2608
2018-01-02	5:34:17	37.36208	126.5484	1975	2018-01-02	5:33:21	37.32816	126.5785	2600
2018-01-02	5:34:26	37.36702	126.544	1850	2018-01-02	5:33:22	37.32886	126.5779	2595
2018-01-02	5:34:37	37.37237	126.5393	1725	2018-01-02	5:33:23	37.32955	126.5773	2590
2018-01-02	5:35:49	37.40938	126.5064	825	2018-01-02	5:33:24	37.33025	126.5767	2585
2018-01-02	5:35:57	37.4125	126.5036	750	2018-01-02	5:33:25	37.33094	126.5761	2580
2018-01-02	5:36:06	37.41696	126.4997	625	2018-01-02	5:33:26	37.33164	126.5754	2575
2018-01-02	5:36:13	37.41998	126.4971	550	2018-01-02	5:33:27	37.33233	126.5748	2570
					2018-01-02	5:33:28	37.33303	126.5742	2565
					2018-01-02	5:33:29	37.33372	126.5736	2560
					2018-01-02	5:33:30	37.33442	126.573	2555
					2018-01-02	5:33:31	37.33511	126.5724	2550

Figure 8. Before(Left) and after(Right) interpolation of domestic aircraft ADS-B data
그림 8. 국내 항공기의 ADS-B 데이터 보간 전(좌)/후(우)

V. 머신러닝 기반 항공 이상상황 분석 모델 성능 실험 및 결과

실험을 위한 국내 공항 데이터로는 2019년 1월 1일부터 12월 31일까지 수집된 국내 공항에 이착륙하는 항공기들에 대한 ADS-B 데이터를 사용하였다. 수집된 항공 데이터는 해외 항공 ADS-B 데이터와 유사한 환경을 만들기 위해 데이터 전처리를 진행하였다. 또 국내 항공안전 데이터의 특성에 맞춰 UOC_D 도 이상 징후로 추가되어 고려됨에 따라 기존 항적 사이의 유사도를 통해 이상 징후를 분석 및 탐지하는 모델 사용이 어려워졌다. UOC_D 이상 징후의 항적이 정상 항적 및 Go-Around 와 매우 유사한 항적들이 존재하기 때문이다. Go-around 와 UOC_D 이상 징후를 탐지하기 위해 변경된 항공기 이상 징후 분석 모델에 적용하기 위해 시계열 항공 데이터를 특징 데이터로 치환하는 데이터 전처리를 진행하였다. 이후 각 모델에 대해 항공기 이상 징후 분석 결과를 정확도, 정밀도, 재현성, F-score 를 도출하여 성능을 비교하고 최적의 국내 공항 항공기 이상 징후 분석 모델을 확인하였다.

5.1 항공기 이상 징후 분석을 위한 머신러닝 모델 후보군 및 선정

해외에서 연구한 항공기 이상 징후 분석 모델은 회귀 분석 모델로 RF[8], SVM[9], Logistic, XGBoost 등 모델을 사용하였다. 이러한 회귀 분석 모델은 이상 징후인 Go-around 와 정상 항로만을 구분하기 위해 사용한 모델이다. 국내 항공 환경에서의 이상 징후는 Go-around 와 더불어 UOC_D 까지 구분을 해야 한다. 이러한 국내 항공 환경에서의 이상 징후를 분석하기

위해 본 논문에서는 K-means[10], GMM[11], 및 DBSCAN[12] 클러스터링 모델을 적용하여 기존의 회귀 분석 모델과 결과를 비교하고자 한다.

5.2 머신러닝 분석 모델을 통한 항공기 이상 징후 분석

interpolation 된 항공 데이터를 기반으로 착륙 단계에서의 항적 데이터를 Fig. 9 과 같이 시간과 고도 데이터로 확인하였다. 이상 징후의 항공 데이터는 착륙 단계에서 고도가 상승하는 것을 확인하였다. 고도 상승의 형태는 일정하지 않지만, 일정 시간 동안 고도가 급격히 상승하는 것을 알 수 있다. 고도 상승 지점을 선정하여 회귀 분석 모델을 적용하기 위해 고도 상승 최고점을 TFR(Track From Runway) 거리를 기준으로 데이터 분포도를 Figure. 10 과 같이 확인하였다.

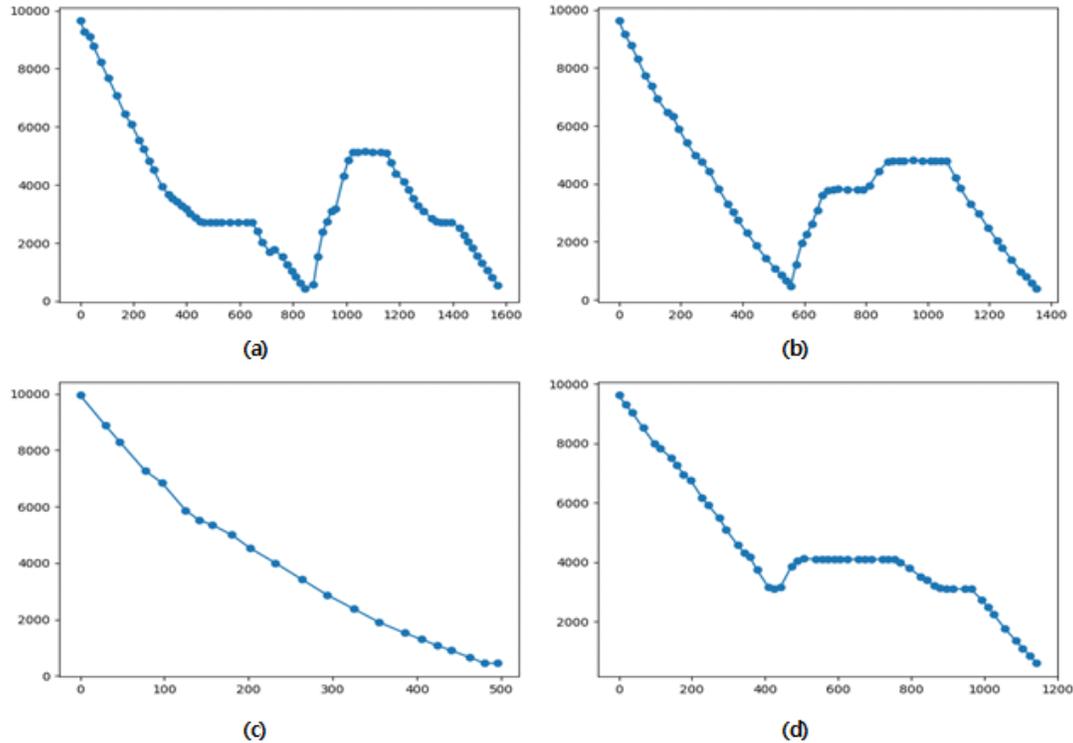


Figure 9. Altitude graph over time (a), (b) is Go-Around, (c) is normal data, (d) is UOC-D data
그림 9. 시간에 따른 고도 그래프 (a), (b)는 Go-Around, (c)는 Normal data, (d)는 UOC-D data

Figure. 10 에 따르면 최고 고도 상승 지점은 9.75~10.92km 에 가장 많이 분포되어 있으며 12.10~13.28km 가 두 번째로 많이 분포되어 있다. 여기에서 추가로 확인해야 할 데이터는 16.82km 이후의 데이터이다. 16.82km 이전의 고도 데이터는 점차 줄어드는 것을 확인하였으나 16.82km 이후 데이터가 일부 존재하는 것을 확인할 수 있으며 27km 이후에도 일부 최고 고도 데이터가 존재하는 것이 확인된다. 이는 국내 항공기 이상 징후의 특징으로 이상 징후 발현 후 다른 활주로 및 인접한 다른 공항으로 착륙을 하는 경우이다. 국내 공항의 경우 인접한 공항이 존재하여 이러한 형태가 보이지만 해외 항공 데이터에서는 이러한 데이터는 잘 존재하지 않는다. 최고 고도 데이터를 기준으로 지정한 TFR 를 기준으로 RF, SVM, Logistic, XGBoost 등의 회귀 분석 모델을 통한 결과 확인하였다.

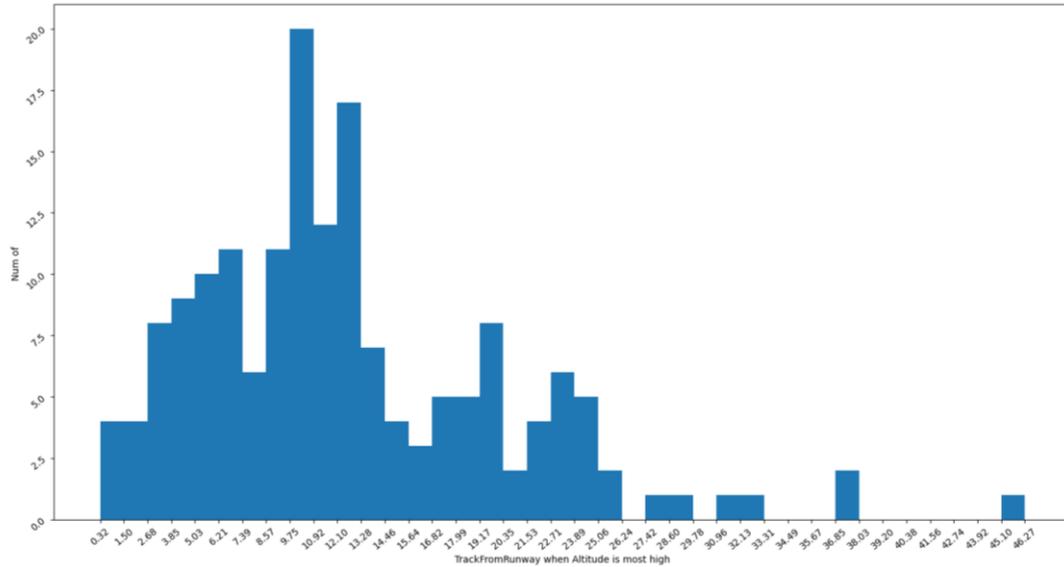


Figure 10. Distribution of highest altitudes by TFR(Track From Runway)

그림 10. TFR(Track From Runway)에 따른 최고 고도 분포

Table 2 는 해외 항공 데이터를 TFR 를 회귀 분석 모델로 분석한 기존 연구의 성능 결과이다. 해외 항공 데이터는 정확도와 정밀도 측면에서 XGBoost 가 가장 높았고, 재현율 및 F-score 부분에서는 SVM 이 가장 높았다. Table 3 은 국내 항공 데이터를 interpolation 진행하여 기존 연구에 사용한 데이터와 유사한 데이터 형태로 가공하고 회귀 분석 모델을 적용한 성능 결과표이다. 정확도는 Logistic 과 XGBoost 가 가장 높았으며 정밀도, 재현율, F-score 전부 Logistic 모델이 가장 높은 결과를 보인다. Figure. 11 은 국내 항공 데이터를 기반으로 회귀 분석 모델을 적용한 결과를 시각화한 그래프이다.

Table 2. Overseas Aircraft Anomaly Regression Analysis Model Performance Results

표 2. 해외 항공기 이상 감지 회귀 분석 모델의 성능 결과

	Accuracy	Precision	Recall	F2-Score
Logistic	0.825	0.313	0.757	0.589
XGBoost	0.885	0.425	0.696	0.617
RF	0.863	0.374	0.732	0.615
SVM	0.870	0.395	0.778	0.652

Table 3. Domestic Aircraft Anomaly Regression Analysis Model Performance Results

표 3. 국내 항공기 이상 감지 회귀 분석 모델의 성능 결과

	Accuracy	Precision	Recall	F2-Score
Logistic	0.8	0.835	0.682	0.697
XGBoost	0.8	0.512	0.549	0.536
RF	0.794	0.545	0.529	0.532
SVM	0.717	0.449	0.444	0.445

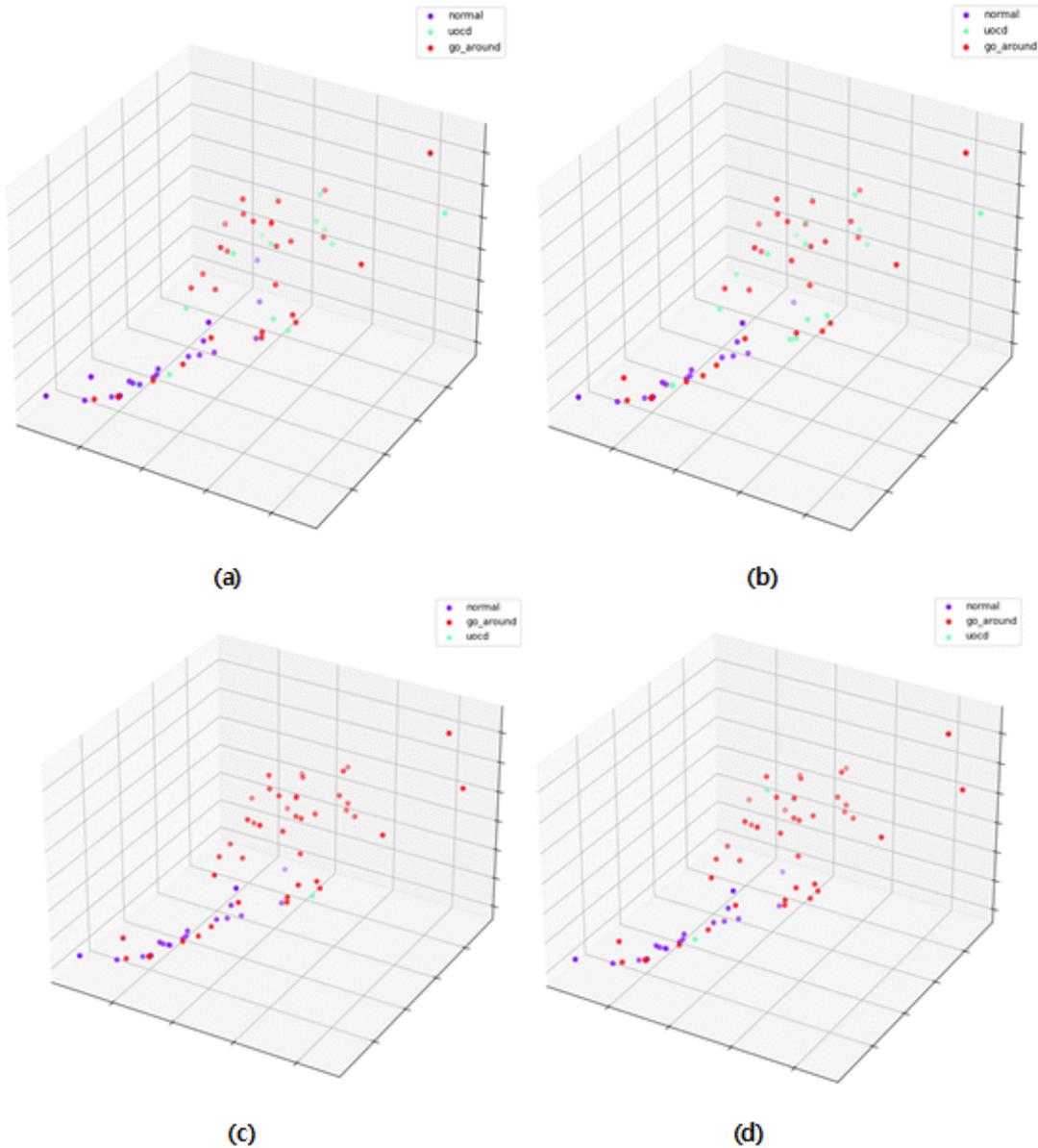


Figure 11. Regression analysis model result graph. (a) XGBoost, (b) RF, (c) SVM, (d) Logistic
그림 11. 회귀 분석 모델 결과 그래프 (a) XGBoost, (b) RF, (c) SVM, (d) Logistic

해외 데이터를 기반으로 나온 분석 모델 성능과 국내 데이터를 기반으로 나온 분석 모델 성능을 살펴보면 국내 데이터를 기반으로 나온 결과가 해외 데이터를 사용한 것에 비해 조금 낮게 나온 것을 알 수 있다. 특히 기존 연구에서는 XGBoost 모델이 전체적으로 가장 좋은 결과를 보였으나, 국내 항공 데이터에서는 Logistic 모델이 가장 좋은 성능 결과를 보인다. 전체적인 모델 성능이 낮은 것은 해외 데이터와 비교하면 국내 데이터의 수가 적어 성능이 떨어지는 것으로 보인다. 하지만 기존 XGBoost 모델과 비교하여 Logistic 모델 성능이 더 높게 나타난 것은 국내 항공기 데이터의 특성으로 확인된다. Figure. 12는 Logistic 모델의 군집 결과 항적 가시화 그래프이다.

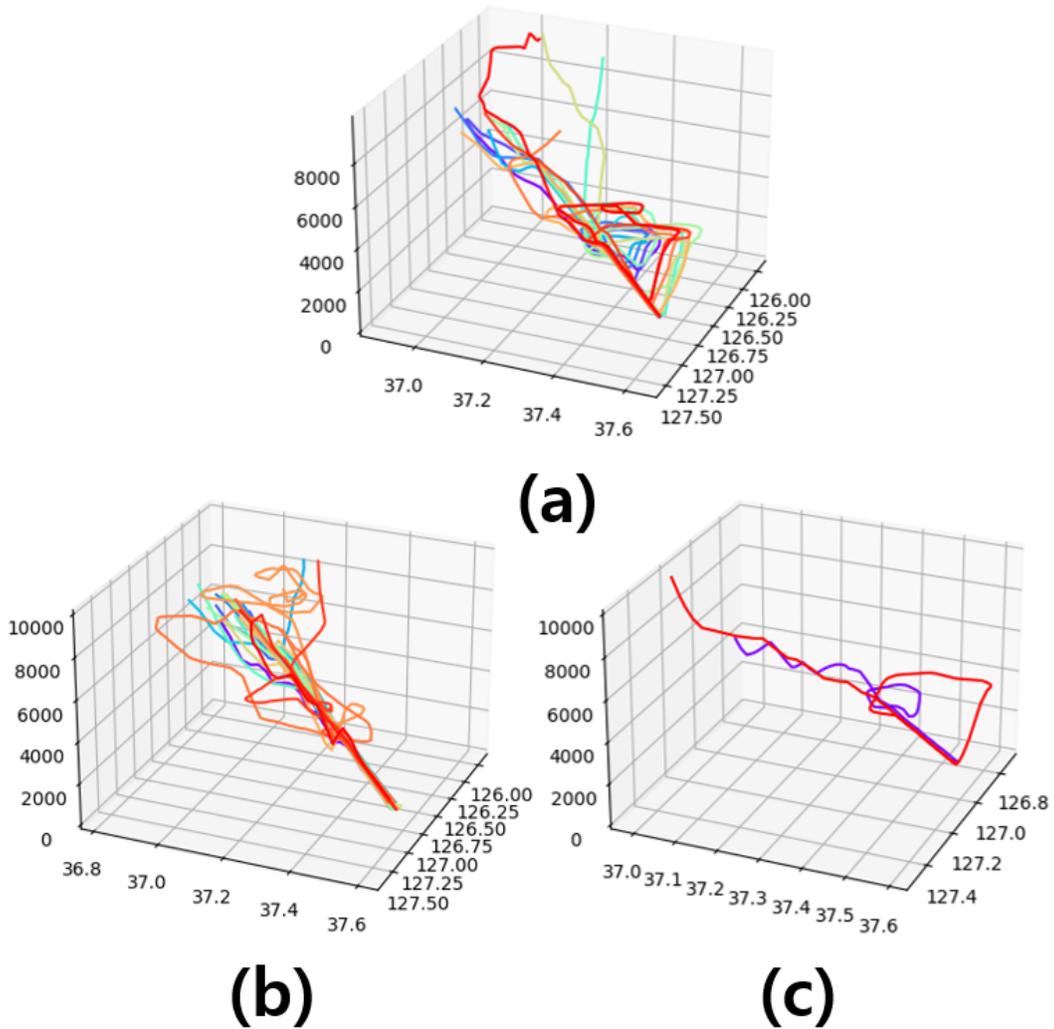


Figure 12. Logistic regression result trajectory visualization graph (a) Go-Around, (b) Normal, (c) UOC-D
 그림 12. 로지스틱 회귀 결과 경로 시각화 그래프 (a) Go-Around, (b) Normal, (c) UOC-D

항공기 이상 징후 클러스터링 모델은 데이터 전처리가 완료된 항공 데이터를 기반으로 분석하여 결과를 확인하였다. Figure. 13 는 rule-base 로 이상 징후를 분류한 결과이다. 그래프에서 노란색은 normal, 보라색은 UOC_D, 초록색은 Go-around 이다. UOC_D 데이터는 Go-around 와 normal 과 유사한 영역에 존재하는 데이터가 다수 확인된다.

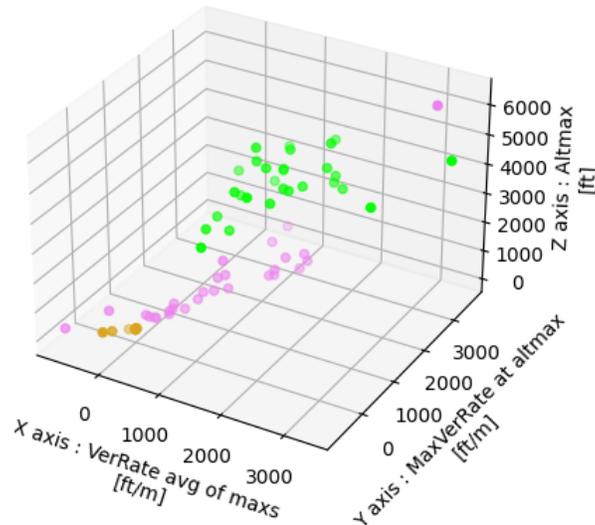


Figure 13. Aircraft Anomaly Classification Results
그림 13. 항공기 이상 분류 결과

Table 4 는 국내 항공 데이터를 클러스터링 분석 모델로 분석한 모델의 성능 결과이다. 또한 Figure. 14, 15, 16 은 국내 항공 데이터를 기반으로 GMM, K-means, DBSCAN 모델을 적용한 결과를 시각화한 그래프이다. 전체적인 모델의 성능 결과를 회귀 분석 모델과 비교하여 높은 결과를 보이며 이는 해외 데이터를 통해 나온 결과와 비교하여도 높은 성능 결과를 보인다.

Table 4. Domestic Aircraft Anomaly Clustering Model Performance Results
표 4. 국내 항공기 이상 감지 회귀 분석 모델의 성능 결과

	Accuracy	Precision	Recall	F2-Score
GMM	0.889	0.773	0.895	0.867
K-Means	0.857	1	0.526	0.581
DBSCAN	0.778	1	0.263	0.308

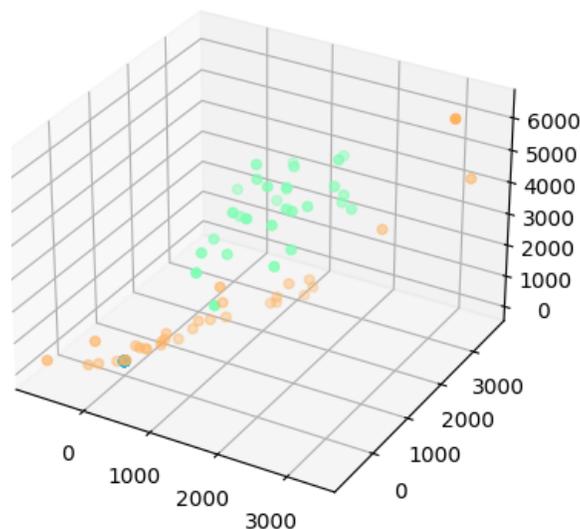


Figure 14. GMM Clustering Result
그림 14. GMM 클러스터링 결과

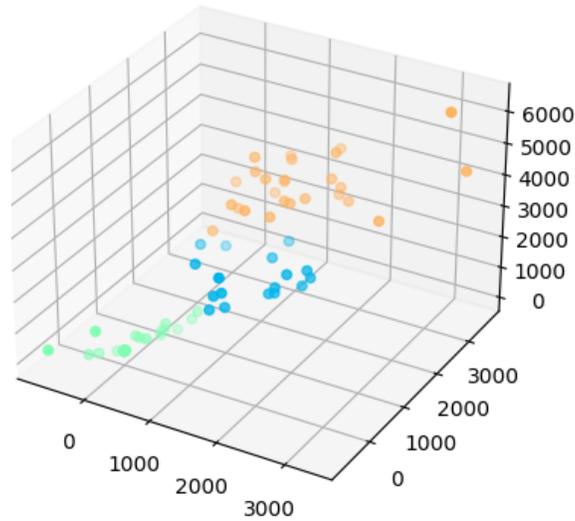


Figure 15. K-Means Clustering Result
그림 15. K-Means 클러스터링 결과

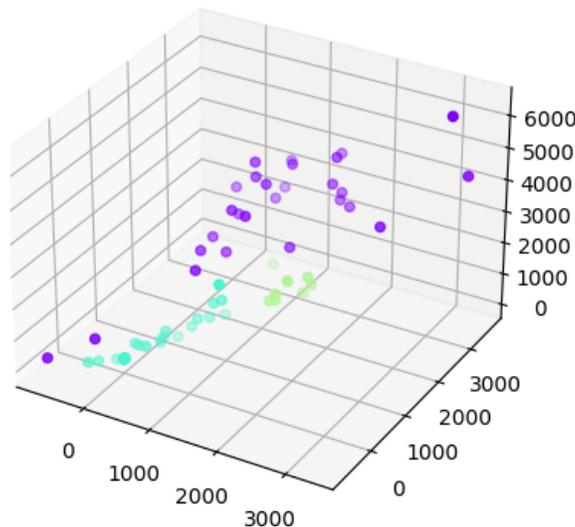


Figure 16. DBSCAN Clustering Result
그림 16. DBSCAN 클러스터링 결과

GMM 클러스터링은 선정된 클러스터의 개수에 맞춰 가우시안 분포를 가진 데이터로 군집을 찾는 방식이다. 본 연구에서 활용한 특징 데이터는 각 이상 징후가 정규 분포와 유사한 분포의 형태를 가지고 있어 정규 분포를 찾아 군집을 나누는 GMM 클러스터링 결과가 이상 징후를 잘 구분하였다. 하지만 전체 데이터의 수가 부족하여 정규 분포의 중심으로부터 멀어질수록 오차가 발생하는 것을 확인할 수 있다.

K-means 클러스터링은 각 군집 중심에서부터 가까운 데이터를 찾는 방식으로 이상 징후별 중심은 유사하게 찾았지만 각 이상 징후가 군집 중심으로부터 거리가 가깝게 집중되지 않고 개별적인 분포를 가진 형태로 실제 이상 징후와 다른 오차 데이터가 다소 많이 발견되었다. 하지만 normal 데이터는 매우 근접한 거리에 집중되어 군집을 잘 이루어 예측이 잘 진행되었기 때문에, 모델 성능 결과가 높은 정확도를 보인다.

DBSCAN 클러스터링은 앞선 클러스터링과 다른 밀도 기반의 클러스터링으로 각 데이터 간의 거리가 가까운 데이터들로 자동화 군집을 이루는 방식으로 군집의 개수를 모델이 결정한다.

클러스터링 결과 대부분의 UOC_D 데이터가 다른 군집으로 흡수되어 가장 낮은 모델 성능 결과를 보인다.

Figure. 17 은 GMM 클러스터링 결과 항적 가시화 그래프이다. 각 군집의 항적은 고도 상승량이 명확한 차이를 보인다. 이러한 차이를 통해 각 이상 징후를 시각적으로 구분할 수 있다.

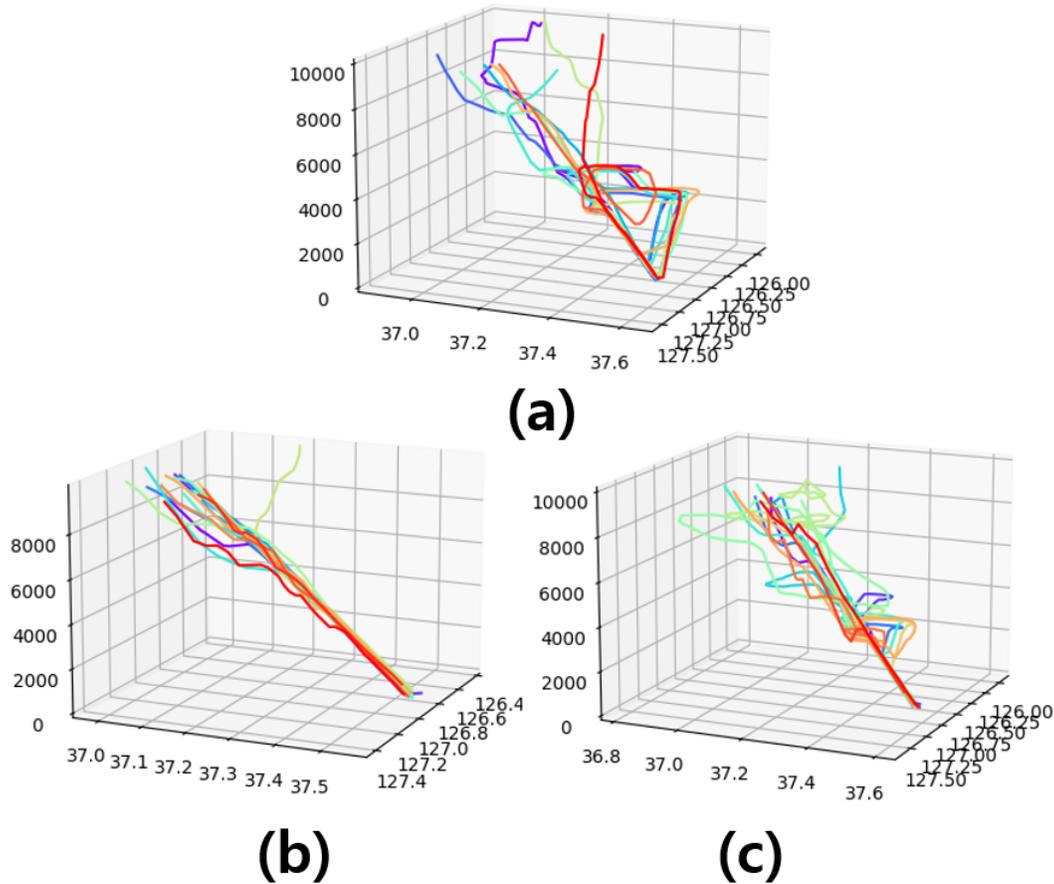


Figure 17. GMM Clustering result trajectory visualization graph (a) Go-Around, (b) Normal, (c) UOC-D
그림 17. GMM 클러스터링 결과 경로 시각화 그래프 (a) Go-Around, (b) Normal, (c) UOC-D

회귀분석 모델과 클러스터링 모델의 성능을 비교해 보았을 때 가장 높은 성능을 보인 것은 GMM 으로 나타난다. GMM 을 제외한 모델의 성능은 순차적으로 K-means, Logistic 순으로 높은 결과를 보이며 가장 낮은 결과는 DBSCAN 으로 확인된다. 이를 통해 국내 항공기 환경에서의 이상 징후 탐지에서는 GMM 이 가장 최적화된 분석 모델인 것을 확인할 수 있다.

VI. 결론

본 논문에서는 국내 항공 데이터를 기반으로 클러스터링 모델 및 회귀 분석 모델을 적용해 항공기 이상 징후를 구분하는 모델 성능을 정확도, 정밀도, 재현율, F-score 를 통해 평가하였다. 기존 해외 연구에서는 회귀 분석 모델 중 RF 모델이 항공기 이상 징후를 분류하는 가장 높은 성능을 보였으나 국내 데이터에서는 GMM 클러스터링 모델이 가장 높은 성능을 보였다. 이는 해외 항공 데이터와 다른 국내 항공 데이터의 특성으로 본 연구를 통해 국내에 최적화된 항공기 이상 징후 분석 모델을 제시하였다.

다만, 이러한 항공 데이터들의 분석을 통해 국내 데이터를 활용하여 항공 데이터의 이상 징후를 탐지할 수는 있으나 실시간으로 수집되는 개별 항공 데이터를 기반으로 이상 징후의 발생 여부를 예측하는 것은 한계점으로 제시되고 있다. 본 논문은 국내 항공 이상 징후의 발생을 분석하는 연구의 기초 연구로써, 최종적으로는 이상 징후의 발생 여부를 실시간으로 예측하는 것이 목표이다. 향후에는 본 연구에서 분류된 이상 징후 데이터를 분석하여 이상 징후가 발생하기 전 나타나는 특징을 기반으로 LSTM 기반의 딥러닝 모델을 적용하여 시계열 항공 궤적 데이터의 이상 징후 발생 여부를 예측하는 연구가 진행되어야 할 것이다.

VII. Acknowledge

본 연구는 국토교통부 항공선진화사업의 일환으로, "빅데이터 기반 항공안전관리 기술개발 및 플랫폼 구축"사업의 4 단계(2023 년도) 연구를 통해 수행되었습니다.

VIII. 참고문헌

- [1] X. Zhao, Y. Rao, J. Cai, W. Ma, "Abnormal trajectory detection based on a sparse subgraph", IEEE Access, Vol. 8, 2020, pp. 29987-30000
- [2] Z. Shi, M. Xu, Q. Pan, B. Yan, and H. Zhang, "LSTM-based Flight Trajectory Prediction", 2018 International Joint Conference on Neural Networks, 2018, pp. 1-8
- [3] L. Dai, Y. Liu, and M. Hansen. "Predicting Go-around Occurrence with Input-Output Hidden Markov Model." International Conference on Research in Air Transportation, 2020
- [4] S.R. Proud, "Go-around detection using crowd-sourced ADS-B position data." Aerospace, Vol. 7, No. 2, 2020, pp. 16
- [5] J.H.Cho, H.J.Baik, "A Study on the Characteristics of Low-Level Wind Shear at Jeju International Airport from Go-Around Flight Perspective", Journal of the Korean Society for Aviation and Aeronautics, Vol. 29, No.1, 2021, pp. 1-8
- [6] J.Zhang, L.I.U.Wei, Z.H.U.Yanbo, "Study of ADS-B data evaluation.", Chinese Journal of Aeronautics, Vol. 24, No. 4, 2011, pp. 461-466
- [7] M.Schafer, M.Strohmeier, V.Lenders, "Bringing up OpenSky: A large-scale ADS-B sensor network for research.", IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks. IEEE, 2014, pp. 83-94
- [8] Rigatti, Steven J. "Random forest.", Journal of Insurance Medicine, Vol. 47, No.1, 2017, pp. 31-39
- [9] W.S. Noble, "What is a support vector machine?", Nature biotechnology, Vol. 24, No. 12, 2006, pp. 1565-1567
- [10] H.H. Bock, "Clustering Methods: A History of k-Means Algorithms", Selected Contributions in data analysis and classification, 2007, pp. 161-172
- [11] E. Alpayd, "Soft vector quantization and the EM Algorithm", neural Network, Vol. 11, Issue. 3, 1998, pp. 467-477
- [12] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in Proceeding of the ACM International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226-231,

저자소개



이상호 (Sang Ho Lee)

2012년 05월 Digipen Institute of Technology Computer Science 학사
 2023년 08월 연세대학교 글로벌융합공학부 공학박사
 2021년 06월~현재 주식회사 애나 기술이사(CTO)

관심분야: 인공지능, 메타버스, 블록체인



손일락 (Ilrak Son)

2023년 2월 인하대학교 기계공학과 학사
 2022년 04월 주식회사 애나 선임연구원

관심분야: 인공지능, GAN



정규호 (Kyuho Jeong)

2020년 8월 명지대학교 컴퓨터공학과 학사
 2022년 05월~현재 주식회사 애나 연구원

관심분야: 인공지능, 프로그래밍



박노삼 (Nohsam Park)

1999년 2월 경북대학교 대학원 컴퓨터공학과 석사
 2013년 12월 Carnegie Mellon University Software Engineering 석사
 2001년 1월~현재 한국전자통신연구원 책임연구원

관심분야: 인공지능, 항공 안전, 디지털 융합
