

기상정보를 활용한 북한 탄도미사일 발사 가능 환경 예측 모델 연구

A Study on Predicting the Readiness Posture Environment of North Korea's Ballistic Missile using Weather Information

조성일¹⁾ · 이춘주²⁾ · 문호석³⁾

Sungeel Jo · Choonjoo Lee · Hoseok Moon

ABSTRACT

Existing research on North Korea's ballistic missile launches has focused on follow-up management of pattern analysis linking ballistic missiles and nuclear development, as well as tracking and predicting the range of ballistic missile orbits. Most of the predictions related to North Korea's ballistic missile launch were viewed from a political and security perspective related to North Korea. These existing studies can also be seen as dealing with the intentions of the North Korean command in light of various situations. However, even if North Korea intends to launch a ballistic missile, it must also have a weather environment to launch it.

Therefore, this study proposes a model to predict the weather environment of North Korean ballistic missiles using past North Korean weather information and ballistic missile launch information. As a result of the study, random forest and adaboost models among classification models show high prediction accuracy of over 94%. Using the results from a decision tree model to interpret key weather variables that significantly influence ballistic missile launch predictions, it was shown that conditions are unfavorable for missile launch when the wind speed is 2.5m/s or greater. This is in line with the fact that too strong ground winds can negatively impact rocket launches. Such a model, trained on past weather data, can be used to predict suitable weather conditions for ballistic missile launches by inputting new weather data.

Key Words : North Korea's Ballistic Missile, Readiness Posture Environment Prediction models, Weather Information, Machine Learning, Cross Validation, National security

논문접수일 : 2023년 4월 28일, 심사일 : 2023년 5월 31일, 게재확정일 : 2023년 6월 7일

1) 국방대학교 국방과학학과 석사과정

2) 국방대학교 무기체계전공 교수

3) 국방대학교 국방과학학과 교수 / 교신저자(corresponding author)

1. 서론

북한의 탄도미사일 발사 여부를 결정하는 것에서 제일 중요한 점은 ‘북한의 의도’이다. 다시 말해서 북한 최고 지도자의 의도에 따라서 발사 여부가 결정될 것이다. 그러나 분명한 ‘발사 의도’가 있더라도 탄도미사일 발사 환경이 조성되어있지 않으면 실행할 수 없다.

본 연구는 북한의 탄도미사일 발사 여부에 관한 예측모델링에 관한 연구이다. 본 연구에서의 예측은 ‘북한 지휘부의 의도’를 예측하는 것은 아니며, 북한이 탄도미사일을 발사할 수 있는 가능한 환경이 조성되었는지를 북한의 기상정보를 통해서 예측하려고 한다. 기상적인 측면에서 발사 환경이 안 되어 있으면 아무리 의도가 있어도 미사일 발사는 불가능하기 때문이다. 본 연구는 북한의 기상정보를 이용해서 북한의 탄도미사일 발사 환경을 예측하려는 측면에 초점을 둔 첫 연구로서 의의가 있다.

북한 탄도미사일의 역사는 1968년 구소련이 제공한 Frog 로켓으로부터 출발한다. 북한은 한미의 우세한 공군 전력을 극복하고 전략적 우위를 확보하고자 탄도미사일 개발에 착수하였다. 1980년대 스커드-B(사거리 300 ~ 500km) 미사일을 개발했고 1990년대 들어선 스커드-C로 사거리를 500km 연장했다[1]. 또한 1990년대 초부터 2,000~2,500km 및 4,000~6,000km의 장거리 탄도미사일인 대포동미사일 1, 2를 개발하기 시작하였다[2]. 2012년부터 신형 액체·고체 추진 탄도미사일에 대한 연구개발을 추진 하여 2016년 개발에 성공한 화성-12형으로 3회에 걸쳐 일본 상공을 통과하는 시험 발사를 하는 등 향후에도 핵 무력 고도화를 위한 미사일 개발과 시험 발사를 지속할 것으로 예상된다[3].

과거 시기별(1984 ~ 2017) 북한 탄도미사일·핵 관련 도발 횟수를 분석해보면 대부분 매년

5월 주변으로 많이 일어난 것을 알 수 있으며, 김일성 시기 8회 중 4회가 5월에 발생했고 김정일 시기 역시 28건 중 5월이 8건으로 가장 많았다. 김정은 집권 이후 또한 2017년까지 전체 76회의 도발 중 42회(55%)가 3 ~ 5월 사이에 일어나며 유사한 결과를 보여주었는데[4] 이를 통해 북한이 계절적 요소를 고려하여 핵 및 미사일 개발을 추진한다고 유추할 수 있다. 즉 ‘북한의 지리적 특성과 겨울철에 매우 낮은 기온 등의 환경 요소를 고려한 것’이라고 볼 수 있다.

하지만, 2022년도에 북한은 이전과는 완전히 다른 탄도미사일 패턴을 보여주었다. <표 1-1>은 '22년 월별 북한의 탄도미사일 도발 현황을 보여주는데 10월과 11월에 각각 6회로 윤석열 정부 출범 후에는 우리나라와 미군의 훈련간 많은 탄도미사일을 발사했음을 알 수 있다.

<표 1-1> 2022년 탄도미사일 도발 일지

구분	횟수[3]	비 고
1월	6	한반도 정세 주도권 확보 조선노동당 정치국 회의 (1.19.)
2월	1	대미 협상력 제고 김정일 생일(2.16.)
3월	3	김정은 국가 우주개발국 현지지도(3.10.)
4월	1	한미 연합 지휘소훈련 (4.18.~26.)
5월	4	윤석열 정부 취임(5.10.) 한미 정상회담(5.21.)
6월	1	한미 해군간 항모강습단 연합훈련(6.2.~4.)
9월	3	한미 연합 해상 훈련 (9.26~29)
10월	6	노동당 창건일(10.10.), 호국 훈련(10.17.~28.)
11월	6	한미 공중 훈련 비질런트 스톱(10.31.~11.5.)
12월	3	유엔총회 북한인권결의안 채택(12.15.)
계	34	-

2022년은 북한이 역대 가장 많은 탄도미사일 도발을 실시한 해로 총 34회를 시행했다[3]. 과거 어떤 시기에 비해 도발 횟수가 매우 많아졌는데, 특히 3월 ~ 5월에 집중되었던 과거와는 다르게[4] 1·10·11월에 6회나 도발을 시행하는 등 다양한 도발 패턴을 보여주었다. 이러한 상황에서는 새로운 도발 패턴에 대한 대비와 더불어 점점 더 발전된 미사일의 위협에 대한 대비가 필요하다. 따라서 적의 새로운 도발 패턴에 대응하고 점점 고도화되는 북한 탄도미사일의 위협에 대응하기 위해 기상청의 기상 데이터를 활용하여 북한 탄도미사일 발사 환경이 기상 환경측면에서 준비되어 있는지를 예측해 보았다. 이를 위해서 과거 북한의 탄도미사일 발사일 기상과 도발을 하지 않은 날의 기상 데이터를 학습데이터로 활용하여 기상적으로 탄도미사일 도발 환경이 준비되어 있는지를 예측하려고 한다.

2. 관련 연구

2.1 도발 패턴에 관한 연구

북한의 미사일 실험 및 발사 진행 절차를 포괄적 전략(분석틀) 측면에서 보면 주변 정세 파악, 해당 담당자 토의, 실험/발사 계획수립/보고, 김정은 최종결정 순으로 진행된다[4]. 북한은 도발 전 먼저 한국과 미국의 정세를 세부적으로 파악한다[5]. 이후 이해관계를 고려한 실험, 발사 수단 및 방법토의를 시행하는데 이때 관련 내용은 독대로 진행되기 때문에 누구도 관련 내용을 알 수가 없다[4]. 북한이 우리나라와 세계에 가장 위협을 주는 분야가 핵이다. 핵과 관련해서 북한은 핵실험에 앞서 장거리 탄도미사일을 먼저 발사하고 이후 핵실험 예고, 외무성 성명 발표, UN 제재에 대한 반박 성명순으로 반복적인 패턴을 보여 오고 있다. 이러한 북한의 핵실험 관련 도발 패턴을 볼 때 탄도미사일 발사 여부를 예측하는 것은 북한

핵실험에도 대응할 수 있다는 의미가 있다.

2.2 탄도 미사일의 발사지점 예측 연구

탐지된 속도를 활용한 발사지점 예측 연구를 보면 발사후 Boost 단계와 Ballistic 단계를 구분하여(PDLS : Phase Division Least Square) 발사지점을 예측한다[6]. 탄도미사일 속도와 각도, 진행방향 중 방위각을 이용하여 Least Square 방법으로 이전 속도를 추정하고, 추정된 속도 특성 및 현재위치와 Euler 적분법을 이용하여 발사지점을 예측하여 최초 탐지 시점이 빠를수록 발사 지점에 대한 거리 오차가 줄어드는 성능을 보여주었다. 하지만 종말단계에서 추력이 없을 때 탄도미사일이 탐지된다면 실제 발사장소와 큰 오차를 보이며 발사를 예측하는 경계의 의미보다 이미 발사된 후 원점을 추정한다는 점에서 한계가 있다. 우리에게 실질적 위협인 북한의 탄도미사일을 대비하기위해선 북한이 발사하기 전 징후를 파악하고 이에 대한 조치가 필요하다.

2.3 감시정찰자산 활용 탄도 미사일 발사 탐지 연구

위성 기술을 활용하여 북한의 탄도미사일을 탐지하는 연구를 살펴보면 위성과 지·해상 레이더를 활용한 방안을 제시하고 있다. 그 일환으로 융합센터를 활용한 정지궤도 적외선 위성, 저궤도 SAR 위성, 지·해상 레이더 통합 탐지방안으로 실질적으로 경향을 확인하며 대응하는 방안도 제시되고 있다[7]. 특히 조기경보 위성 기술을 활용하여 북한의 탄도미사일 활동을 감지한다면 탄도미사일 발사 등의 북한위협에 보다 세밀하게 대응할 수 있을 것이다[8]. 하지만 북한은 잠수함이나 터널 등을 이용하며 다양한 발사 수단을 동원하여 탄도미사일을 발사하고 있기 때문에 기존 탐지자산으로 탐지를

더욱 어렵게 하고 있다는 점도 고려해야 한다. 즉, 기존 탐지자산으로는 예측에 한계가 있다.

3. 연구 방법

3.1 연구 대상

본 연구의 목적은 북한의 기상 데이터를 활용하여 탄도미사일 발사 가능한 환경을 예측하는 모델을 개발하는 것으로, 실제 북한이 탄도미사일을 발사한 날 기상을 도발할 수 있는 기상으로 가정하고 분석하였다. 기상정보를 확인하려면 지역 정보가 중요한데, 지역 정보는 현재까지 북한이 탄도미사일을 발사할 수 있는 곳으로 예측되는 북한의 9개 도를 대상으로 하였다. 최종적으로 북한 지역 9개 도 중에 비도발지역(북한이 한번도 도발하지 않은 지역)인 양강도, 함경북도, 황해남도를 제외하고 6개 도 지역(강원도, 자강도, 평안남도, 평안북도, 함경남도(2), 황해북도)의 7개 도시(원산, 강계, 평양, 신의주, 신포, 함흥, 개성)를 탄도미사일 도발 가능 지역으로 선정하였다. 선정된 지역에 대한 '22년 기상자료를 수집하고, 탄도미사일과 관련된 관련 연구를 기초로 기상정보를 분석하여 기상정보 가운데 기상변수로 <표 3-1>과 같이 총 11개 변수를 선정하였다.

<표 3-1> 기상 변수

Variable		Unit
avgTa	평균 기온	℃
avgTd	평균 이슬점온도	℃
avgPa	평균 현지기압	hPa
avgWs	평균 풍속	m/s
sumRn	합계 강수량	mm
avgTca	평균 전운량	(0~10)
maxTa	최고 기온	℃
avgHm	평균 습도	%
avgPs	평균 해면기압	hPa
Lmac	중하층 운량	(0~10)
minTa	최저 기온	℃

11개 변수 중 sumRn(결측률 63.78%)과 Lmac(결측률 32.14%)은 다수의 결측치로 인해 모델 구성 간 제외하였다.

기상 변수 중 이슬점온도는 공기가 일정한 압력과 일정한 수증기 함량에서 냉각될 때 포화되는 온도로 이슬이 맺히는 온도이다[9]. 이슬점온도가 낮을수록 공기 중의 수증기 함량이 적어지며 이는 비례관계를 보여준다. 해면기압은 실제 고도가 아닌 평균해면에서 기압으로 환산한 것으로 관측소기압으로부터 경험적으로 추정한다[10]. 해수면 기준으로 기압을 환산했기 때문에 고도에 상관없이 비교할 수 있는 이점이 있다.

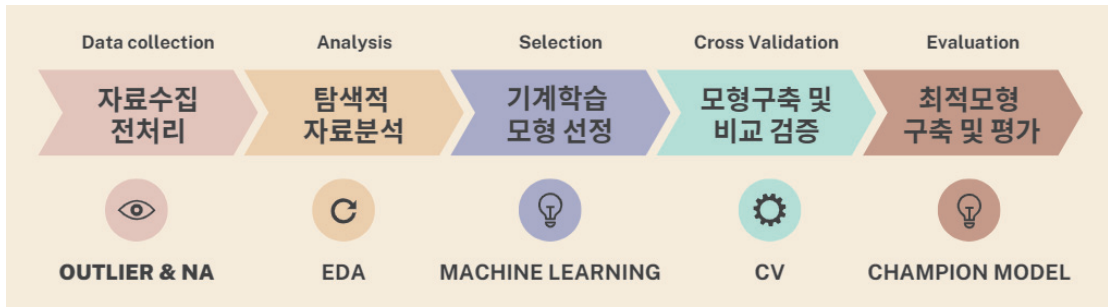
3.2 연구 절차와 방법

북한 탄도미사일 발사 예측 모델을 구성하기 위한 연구 절차는 <그림 3-1>과 같다.

자료수집 및 전처리 단계에서는 북한의 기상 자료를 수집하고 특이값이나 결측치(NA) 등을 확인하는 단계로 자료는 기상청 기상자료개방포털에서 수집하였다⁴⁾. 1973년부터의 북한기상 데이터가 저장되어있으며 이중 가장 도발이 많이 발생했던 해인 2022년의 자료를 수집하였다. 특이값(outlier)은 각 변수별 상자 그림을 통해서 확인하고 제거하였고, 결측치(missing)를 확인하여 변수 평균으로 대체하였다.

탐색적 자료분석(EDA, explanatory data analysis) 단계는 데이터를 있는 그대로 탐색하여 분석하는 것으로 기술 통계기법 등이 주로 사용된다. 잔차, 자료의 재표현, 그래픽을 통한 현시성을 통해 데이터를 있는 그대로 분석할 수 있다. 상관관계 분석간에는 변수간의 선형적인 관계 정도를 측정하여 그 크기와 방향을 파악할 수 있는데 상관관계가 높다면 다중공선성이 발생할 수 있어 변수를 제거하거나 조정하는 등의 과정이 필요하다.

4) 기상청 기상자료개방포털(<https://data.kma.go.kr/data/grnd/selectNkRltnList.do?pgmNo=58>)



<그림 3-1> 연구 절차 순서도

기계학습 모형 선정은 앞서 준비한 데이터를 다양한 기계학습 모형들 중에 최적의 모형을 구축할 수 있도록 선정하는 단계이다. 모형은 단일모형과 앙상블로 구성하였다. 딥러닝은 제외하였는데 이는 딥러닝에 필요한 만큼의 자료 크기가 충분하지 않았고, 앙상블 학습이 딥러닝만큼의 충분한 성능을 보여주기 때문이다 [12]. 모형 구축 간 K-Nearest Neighbors algorithm(최근접 이웃 알고리즘, 이하 KNN), CART (classification and regression tree) 알고리즘 기반 의사결정나무(이하 Rpart), random forest(이하 RF), support vector machine(이하 SVM), linear discriminant analysis(선형판별분석, 이하 LDA), quadratic discriminant analysis(이하 QDA), Adaboost, generalized linear model(이하 GLM)를 사용하였다. 모형 구축을 위해 R programming 패키지를 이용하였으며 KNN을 위해 ‘class’, LDA와 QDA를 위해 ‘MASS’, SVM을 위해 ‘e1071’, RF를 위해 ‘randomForest’, CART를 위해 ‘rpart’, Adaboost를 위해 ‘ada’패키지를 사용하였다.

GLM(Generalized linear model)은 일반적인 선형 회귀(OLS)를 일반화한 모델로 여기에서 사용한 Logistic Regression은 반응변수가 0(비도발)과 1(도발)인 이진변수일 때 사용한다. SVM은 gamma와 cost 두 개의 초모수(Hyper

parameter)가 있으며, gamma는 커널의 폭(너비)을 조정하는 값이고 cost는 과적합을 막는 정도를 지정한다. 최적의 parameter값을 찾기 위해 tune.svm함수를 사용하였으며 best performance의 gamma는 0.1, cost는 100의 값을 구할 수 있었다.

모형구축 및 비교와 검증단계는 교차타당도 평가를 시행하는 단계로 ‘70대 30’과 ‘10-fold’ 교차타당도 방법을 평가에 사용하였다. 첫 번째로 ‘70대 30’은 훈련자료 70%, 검증자료 30%로 데이터를 분할한 후 무작위로 10회 반복해서 실험하는 방법이며, 두 번째로 ‘10-fold’ 교차타당도 방법은 데이터를 10등분하여 90%로 모델을 학습시키고 나머지 10%로 검증하는데 10등분한 훈련용 90%와 검증용 10%를 바꿔가며 10번 시행하여 그 결과를 평균 내어 검증하는 방법이다. 즉, 각 반복에서 얻은 평가 점수의 평균이 최종적인 모델의 정확도로 추정된다 [11].

최적모형 구축 및 평가는 검증 결과에서 선택된 적합한 모델에 대해 검증에 사용된 모든 데이터와 함께 100%의 데이터를 활용하여 모델을 다시 학습시켜 최적의 모델을 만들었다. 최적의 모델은 OOB(out of bag)를 사용해서 정확도를 평가할 수 있고, 이를 정확도로 하는 최적모형으로 제안하였다[12].

4. 연구 결과

4.1 탐색적 자료분석

본 연구에서 사용한 데이터는 2022년 북한의 기상 제원이다. 평균 기온(avgTa), 평균 풍속(avgWs) 등 9개 연속형변수 기술통계량을 비교한 결과, avgHm과 avgPa, avgTd 분산이 각각 264.18, 196.65, 166.39로 측정되었다. 이러한 결과는 해당 변수들의 연간 변화율이 상대적으로 크다는 북한 기상의 특징을 보여준다.

연속형 변수간 선형적인 상관관계를 알아보기 위해 피어슨 상관계수를 산출하였다. avgTa

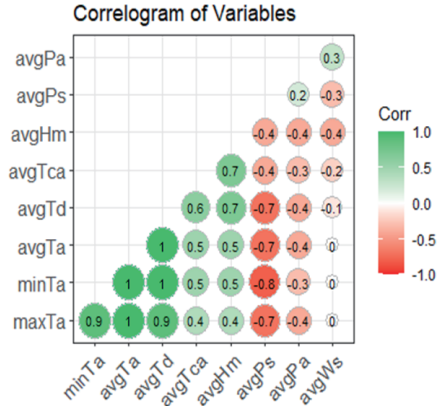
와 avgTd, minTa 등 온도는 상관계수가 1로 강한 선형성을 보여주었다. avgPs와 minTa는 -0.8로 강한 음의 상관관계를 보였고 avgPs와 avgTa, avgTd, maxTa 등 평균 해면기압과 기온간에는 음의 상관관계를 보여주었다. 이는 평균 해면기압이 클수록 기온은 낮아지는 것을 보여주는데 해면기압과 기온이 반비례 관계라는 현상과도 일치한다[13]. 즉, 해수면 공기의 온도가 상승하면 공기가 팽창하기 때문에 밀도가 낮아져 기압이 낮아지게 된다. 반대로 온도가 낮아지면 공기가 수축하게 되며 밀도가 높아져 압력이 증가한다.

<표 4-1> 연속형 변수 기술통계량

변수(단위)	최소	1/4분위수	중앙	평균	3/4분위수	최대	분산
avgTa(°C)	-18.4	1.9	12.2	10.71	20.4	30.5	123.81
avgTd(°C)	-23.5	-5.3	5.3	4.67	15.8	26.4	166.39
avgPa(hPa)	959.6	1001.6	1010.6	1007.8	1018.0	1035.6	196.65
avgWs(m/s)	0	0.8	1.6	1.7	2.4	9	1.46
avgTca(0~10)	0	1.5	5.1	5.01	8.4	10	12.71
maxTa(°C)	-15.9	7.25	18.1	16.43	25.9	37.6	121.39
avgHm(%)	22	58	71	69.19	82	99	264.18
avgPs(hPa)	991.2	1010.8	1018.0	1017.7	1024.6	1043.4	85.35
minTa(°C)	-25.5	-3	7	6.02	16.1	25.9	138.27

<표 4-2> 기상변수간 상관관계 분석

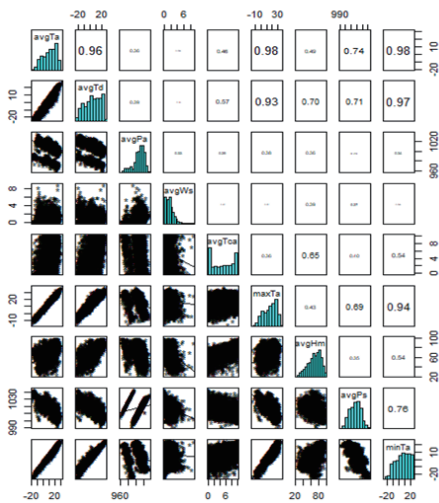
변수	avgTa	avgTd	avgPa	avgWs	avgTca	maxTa	avgHm	avgPs	minTa
avgTa	1	-	-	-	-	-	-	-	-
avgTd	1	1	-	-	-	-	-	-	-
avgPa	-0.4	-0.4	1	-	-	-	-	-	-
avgWs	0	-0.1	0.3	1	-	-	-	-	-
avgTca	0.5	0.6	-0.3	-0.2	1	-	-	-	-
maxTa	1	0.9	-0.4	0	0.4	1	-	-	-
avgHm	0.5	0.7	-0.4	-0.4	0.7	0.4	1	-	-
avgPs	-0.7	-0.7	0.2	-0.3	-0.4	-0.7	-0.4	1	-
minTa	1	1	-0.3	0	0.5	0.9	0.5	-0.8	1



<그림 4-1> 기상변수간 상관관계 분석(1)

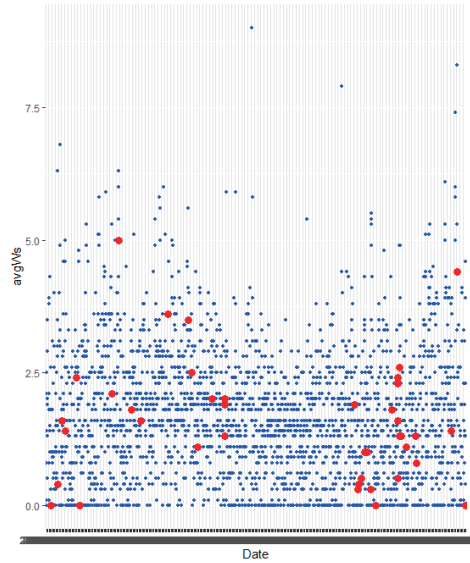
<그림 4-1>과 <그림 4-2>의 상관관계 분석 간 산점도 그래프를 보면 avgTd와 avgTa, avgHm, maxTa, minTa간 강한 선형성을 보여 주고 있다. 이는 평균 이슬점온도가 기온에 영향을 받는다는 것을 유추해볼 수 있다. 기온이 높을수록 대기 중의 수증기 함량이 높아지기 때문에 평균 이슬점 온도도 높아지며, 기온이 낮을수록 대기 중의 수증기 함량이 낮아져 평균 이슬점 온도도 낮아진다. 따라서 이러한 그래프 결과는 평균 이슬점 온도와 기온 간에 밀접한 관계가 있음을 보여준다.

scatter-plot matrix, correlation coef., histogram

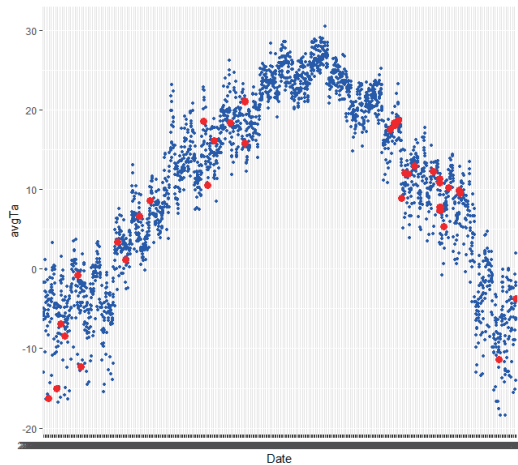


<그림 4-2> 기상변수간 상관관계 분석(2)

<그림 4-3> avgWs 일자별 도발분포 그래프는 연간 일별 평균 풍속을 시각화한 것이다. 파란 점은 비도발 지역이고, 빨간 점은 도발 지역으로 탄도미사일은 대체로 풍속이 0~2.5m/s일 때 발사되었음을 알 수 있다.



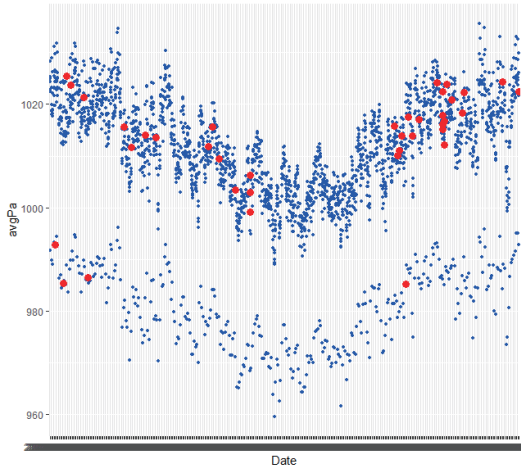
<그림 4-3> avgWs 일자별 도발분포



<그림 4-4> avgTa 일자별 도발분포

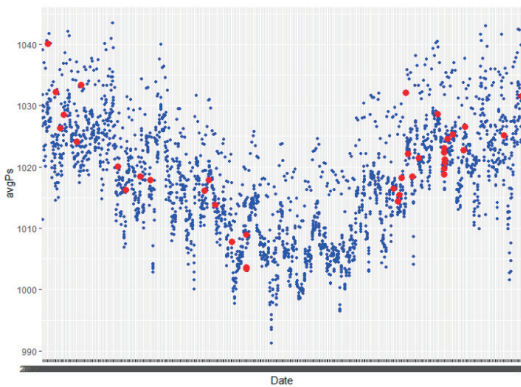
<그림 4-4> avgTa 평균 기온에 따른 발사일을 분석해보면 영하 15도부터 20도까지 발사한 분포를 보여준다. 평균 기온 20도 이상일

때 발사횟수는 1회에 불과하다. 이러한 결과는 기온이 탄도 미사일 발사에 영향을 줄 수 있는 것으로 볼 수도 있으나 과거 사례를 분석해볼 때(2012~2017년 김정은시기 7월 7회, 8월 3회 도발⁴⁾) 다양한 변수에 대한 고려가 필요하다.



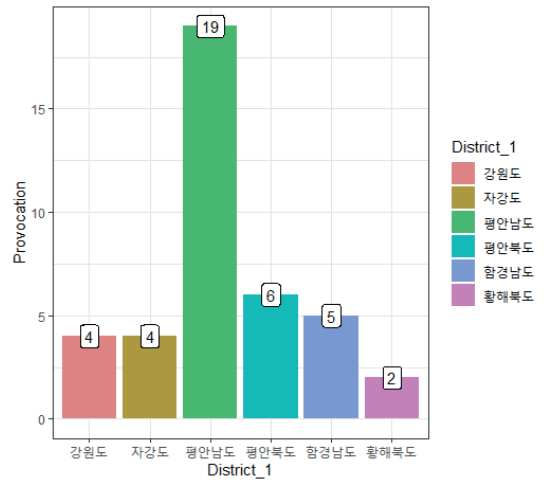
<그림 4-5> avgPa 일자별 도발분포

<그림 4-5>에서 아래 평균현지기압이 낮은(960hpa ~ 995hpa) 지역은 북한 자강도 강계시 지역으로 대체적으로 낮은 현지기압을 보여주었다. 북한 자강도 해발고도가 타 지역보다 높음을 알 수 있는데 평균고도는 해발 803m⁵⁾로 비교적 높은 편에 속한다.



<그림 4-6> avgPs 일자별 도발분포

<그림 4-6> avgPs 평균해면기압은 어떤 관측지점에서 관측된 현지기압을 평균 해수면(해발고도 0m)기준으로 환산하여 산출한 기압이다. 해면기압은 고도가 달라도 비교 가능하여 기상청에서 일기도를 그리는 데 사용된다. 고기압이 형성되는 지역은 대기가 하강하여 공기가 밀집하게 쌓이는 지역이기 때문에 해면기압이 상승하게 된다. 7개 도시 모두 비슷한 형태를 보여주며 대체로 겨울엔 높았으며 여름에는 공기가 뜨거워져서 상승하게 되므로 해면기압이 낮아지는 것을 보여준다.



<그림 4-7> '22년 도발지역별 횟수

도발지역별 도발횟수를 시각화해서 보면 <그림 4-7>과 같이 평안남도(평양 포함)에서 가장 많은 도발이 이루어졌음을 알 수 있다. 주로 순안구역 비행장 일대에서 동해상으로 발사하였으며(15회) 그 외 평안남도 4곳에서 탄도미사일을 발사하였다. 발사체 또한 다양화되고 있는데 평안북도에선 일부 철도 발사형으로 발사하였으며 함경남도 신포 앞바다에선 잠수함에서 발사하는 등 점점 더 예측이 어려워지고 있다.

5) 한국민족문화 대백과사전(<https://encykorea.aks.ac.kr/Article/E0047852>)

4.2 데이터 불균형 해소

SMOTE(Synthetic Minority Over-sampling Technique)는 머신러닝에서 불균형한 데이터를 처리하는 오버샘플링 방법의 하나다. 이 방법은 소수 클래스의 데이터를 무작위로 선택한 후, 그 데이터와 가장 가까운 이웃 데이터들 사이에 새로운 데이터를 합성하는 과정을 반복하여 다수 클래스의 데이터와 균형을 맞춰준다. 이 방법은 합성 데이터를 생성함으로써 샘플의 크기를 증가시키는데, 이를 통해 모델이 불균형 데이터를 처리할 수 있도록 도와준다. 본 연구에서는 1단계로 도발일과 비도발일의 기상을 분석하여 균형을 맞추고 2단계로 SMOTE방법을 적용하였다.

1단계로 도발일과 비도발일의 데이터 불균형이 크기 때문에 불균형데이터를 균형있게 맞춰주는 과정을 거쳤다. 기상으로 탄도미사일 발사 가능한 환경을 예측하기 위한 목적상 각 기상변수의 도발일 기상 평균 $\pm 1\sigma$ (표준편차) 내에 있으면 도발 가능한 것으로 보고 도발일로 가정하였다. 즉, 도발일에 도발하지 않은 지역 이더라도 기상변수별 도발일 기상 평균 $\pm 1\sigma$ 내에 있다면 도발이 가능하다고 보았다.

<표 4-3> Class 수 비교

Category	N-Provocation	Provocation
1단계 (기상 비교 분석)	2493	62
2단계 (SMOTE)	2493	2480

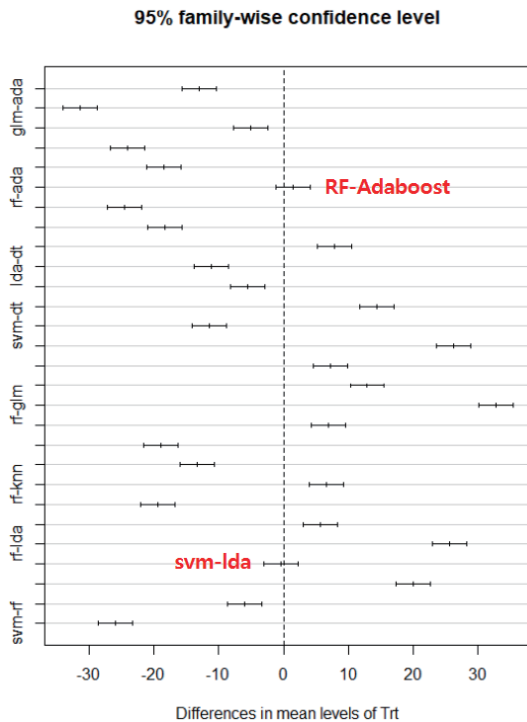
2단계로 SMOTE 방법을 적용하기 전에는 비도발일과 도발일의 데이터 수의 차이가 매우 컸지만, <표 4-3>과 같이 SMOTE를 통해 불균형의 문제를 보완하였다.

4.3 모형구축 및 비교검증

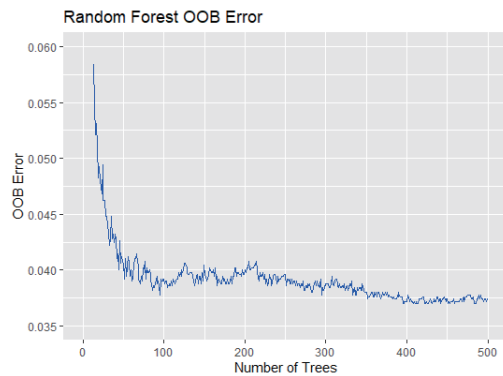
앞서 제시한 8개의 모델로 '70대 30'과 '10-fold' 교차타당도 평가를 하였다. <표4-4> 교차타당도 평가 결과를 보면 'RF'와 'Adaboost'가 가장 좋은 정확도를 보여준다. 결과가 유의미한지 확인하기 위해 10-fold CV결과를 분산분석한 결과 'F(7,72) = 418.1'이고 유의확률(p-value)이 $< 2e-16$ 으로 유의미한 결과를 보여주었다. 즉, 모형간에 정확도에 대한 차이가 있었다. 분산분석 만으로는 어떤 모델결과간에 유의미한 차이가 있는지 알 수 없기 때문에 사후검정을 실시하였다. 이는 분산분석이 집단간의 차이유무만을 말할 뿐 어느 집단간에 차이가 있는지는 알 수 없으며 사후검정을 통해 어느 집단간에 유의미한 차이가 있는지를 확인하기 위해서이다. Tukey의 다중비교를 한 결과 <그림 4-8>과 같이 'RF'와 'Adaboost'는 서로를 제외한 다른 6개 모델과 통계적으로 95% 신뢰수준 하에서 유의미한 차이가 있었다. 'RF'와 'Adaboost'는 Tukey 다중비교 결과 유의확률(p-value)이 0.71로 차이가 있다고 보기 어려웠다.

<표 4-4> 교차타당도 평가 결과

Category		GLM	KNN	Rpart	RF	SVM	LDA	QDA	Adaboost
70:30	평균 (%)	63.41	89.18	82.49	95.58	70.65	71.04	76.34	94.34
	표준편차	1.13	0.93	2.05	0.57	0.99	1.00	1.04	0.67
10-fold	평균 (%)	63.54	89.80	81.92	96.34	70.40	70.82	76.43	94.93
	표준편차	1.52	2.13	2.69	1.08	2.19	2.20	1.69	0.94



<그림 4-8> 모형들의 Tukey 다중비교 결과



<그림 4-9> Random Forest OOB Error

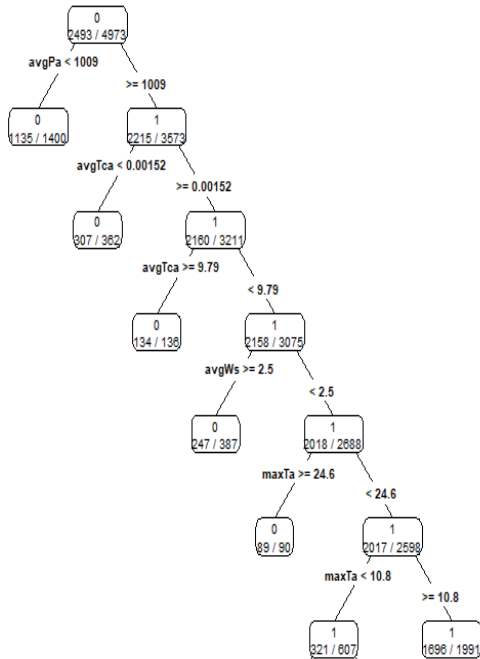
4.4 최적모형 구축 및 평가

교차타당도 평가결과 높은 정확도 및 다른 모델과 비교해서 유의미한 차이를 보여준 앙상블모형인 ‘RF’와 ‘Adaboost’ 모델이 최적의 모형이었다. 해당 모델에 대해 검증데이터를 포함한 100% 데이터를 사용하여 최적모형을 구축하였으며 ‘RF’는 96.26%의 높은 정확도를 보여주었다.

‘RF’의 hyper parameter인 Tree의 수를 구하기 위해 OOB Error를 비교한 결과 400개부터 성능 개선이 미비한 것을 알 수 있다. 트리의 수는 알고리즘의 예측 성능과 모델의 복잡도를 조절하는 중요한 hyper parameter이다. 즉, 높은 예측 성능을 보여주면서 모델의 복잡도가 증가하지 않게 적절한 수의 트리를 설정하는 것이 중요하다. 따라서 적정한 트리의 수는 400으로 결정할 수 있었다

Adaboost는 96.0%의 분류 정확도를 보여주었다. ‘RF’와 0.26% 정확도 차이를 보여주었으나 Tukey 다중비교간 유의미한 차이가 없었으므로 ‘RF’ 성능이 더 좋다고 단정하기 어렵다. 이러한 모델을 통해 과거 기상 데이터를 바탕으로 학습하고 새로운 데이터를 입력하여 탄도 미사일 발사 가능한 기상 환경을 예측하는데 활용할 수 있을 것이다.

4.5 의사결정나무활용 시사점 분석



<그림 4-10> rpart를 활용한 의사결정나무 모형 분류 결과

rpart를 활용한 의사결정나무 모형 분류 결과는 <그림 4-10>과 같다. 주요 분류변수를 해석해보면 첫 번째로, 현지기압이 1009보다 높고 평균전운량이 0.00152~9.79사이일 때 평균풍속이 2.5보다 크거나 같으면 도발이 불가능한 환경이 63.8%(247/387)이었다. 풍속은 실제로 로켓 발사와도 관계가 깊다. 2022년 6월 14일 누리호 2차 발사가 강풍에 의해 연기되었던 것처럼 지상풍이 너무 강하면 발사체 이동 및 설치, 발사 간 위험이 크기 때문에 로켓 발사에 부정적인 영향을 준다고 볼 수 있다.

두 번째로, 첫 번째 환경에서 평균 풍속이 2.5보다 작고 최고 기온이 10.8~24.6도일 때 85.2% (1696/1991)의 탄도미사일 도발 가능한 환경을 보여주었다. 온도가 발사체의 설계 및 운용에 미치는 영향을 보면 연료탱크 등의 밀폐 용기의 내부 압력 변화, 가열로 인한 열 스

트레스, 발사체의 표면 및 내부의 온도 상승으로 인한 기기의 영향등 다양하며 이에 대한 고려 또한 필요하다[14].

5. 결론

본 연구에서는 북한 탄도미사일 발사 가능 기상 환경을 예측하기 위해 북한 기상자료를 활용한 예측 모델을 제시하였다. 본 연구에서는 ‘RF’가 96.34%, ‘Adaboost’가 94.93%의 높은 정확도를 보여주었다.

또한, 탄도미사일 발사 가능 기상 환경 예측에 중요하게 영향을 주는 기상 변수들을 의사결정나무 모형의 결과를 이용해서 해석해 보았다. 해석 결과, 현지기압이 1009보다 높고 평균전운량이 0.00152~9.79 사이면서 풍속이 2.5보다 크거나 같을 때 탄도미사일 발사에 불리한 환경을 보여주었다. 이는 실제 지상풍이 너무 강하면 로켓 발사에 부정적인 영향을 주는 것과 동일한 결과를 보여준다. 반대로 풍속이 2.5보다 작고 최고 기온이 10.8~24.6도일 때 85.2%의 도발 가능한 환경을 보여주었다.

본 연구에서 제안한 탄도미사일 발사 가능 기상환경 예측 모델은 기상 데이터만을 활용하여 분석하였는데, 여러 한계점이 존재한다. 특히, 2022년 북한 기상 데이터만을 사용한 것이기 때문에 다양한 변수들을 추가적으로 수집하여 분석하는 것이 향후에 지속될 필요가 있다.

북한의 탄도미사일 개발은 점점 고도화되고 있으며, 우리나라의 안보에 큰 위협이 되고 있다. 따라서 이를 예측하고 대응하는 노력이 필요하다. 앞으로 본 연구에서 제시한 기상에 대한 분석뿐만 아니라 국·내외 정치적, 군사 외교 및 상황 등과도 연계한 의도 분석이 함께 이루어져야 한다. 이를 통해 우리나라의 안보를 지키고 국민을 보호하는데 기여할 수 있을 것이다.

참 고 문 헌

- [1] 박창권, “북한의 탄도미사일 위협과 한국의 대응체제 발전방향”, 국방정책연구 28(2), 2012
- [2] Yong-Sup Han, “Going Into the ‘unification tunnel’: Strategic and military implications“ in Emerging Threats, Force, Structure, and the Role of Air Power in Korea, edited by Natalie W. Crawford and Chung-In Moon, RAND, 2000.
- [3] 국방부, 「2022 국방백서」, 2022.
- [4] 김경목, 백연주, “북한의 미사일 발사 및 핵실험 패턴에 관한 연구”, 전략연구, 73, 2017.
- [5] 홍민, “김정은 정권 핵·미사일 활동의 주요 특징과 패턴”, 통일연구원 Online Series, 2017
- [6] 김준기, 이동관, 조길석, 송택렬, “단계 분리형 최소 자승법을 이용한 탄도 미사일의 발사지점 예측 연구”, 제어로봇시스템학회, 20(4), 2014
- [7] 김수임, “위성과 지·해상 레이더를 활용한 탄도 미사일 통합 탐지방안 제안 및 효과 분석”, 2021, 아주대학교 석사논문
- [8] 백명진, 최성봉, “한국형 조기경보위성 시스템”, 제어로봇시스템학회, 2012
- [9] 기상청 기상용어사전, Available: <https://data.kma.go.kr/cmmn/dictionaryPopup3.do?wordId=5910>
- [10] 기상청 기상용어사전, Available: <https://data.kma.go.kr/cmmn/dictionaryPopup3.do?wordId=1663>
- [11] 문호석, “기계학습을 이용한 위게임 모델의 근접전투 전장상황 평가를 위한 전문가시스템 연구”, 한국군사학논집, 74(3), 2018.
- [12] 박유성, 『파이썬을 이용한 통계적 머신러닝』, 자유아카데미, 2022.
- [13] 류시찬, “동아시아 해면기압과 한반도 겨울철 기온의 관계”, 2004, 강릉대학교 석사논문
- [14] 김준, 국방재, 문경정, 이재화, 구자호, 박상서, 이효근, “외나로도 우주센터의 기상특성”, 한국항공우주연구원, 2009.
- [15] 김은영, “북한 미사일 주요 일지 및 발표문”. 북한경제리뷰, 3, 2009.
- [16] 정규수, 『ICBM 그리고 한반도』, 지성사, 2012.
- [17] 정규수, 『ICBM, 악마의 유혹』, 지성사, 2012.

저 자 소 개



조성일(E-mail: enjoy_all@naver.com)

2016 육군3사관학교 전자공학 학사
현재 국방대학교 군사운영분석전공 석사과정
관심분야 : 국방 빅데이터·AI, 텍스트 마이닝



문호석(E-mail: hsmoon0329@kndu.ac.kr)

1994 육군사관학교 화학과 학사
2003 고려대학교 전자공학 석사
2006 고려대학교 산업공학 박사
2010 고려대학교 통계학 박사
현재 국방대학교 국방과학학과 교수
국가안전보장문제연구소
군사과학연구센터장
관심분야 : KCTC 빅데이터분석, 데이터과학,
국방 AI, M&S



이춘주(E-mail: sarang90@korea.kr)

1991 육군사관학교 졸업(이학사)
1997 미국 UC Berkeley 핵공학 석사
2005 KDI 국제정책대학원 정책학 석사
2006 서울대학교 기술정책학 박사
현재 국방대학교 무기체계전공 교수
관심분야 : 기술정책, 국방로봇(무인전력),
방위산업정책, 생산성분석,
핵심인프라방호

