

데이터 마이닝 기법을 활용한 군용 항공기 비행 예측모형 및 비행규칙 도출 연구

유경열* · 문영주** · 정대율***

<목 차>

- | | |
|---------------------|--------------------------|
| I. 서론 | 3.3. 연구변수의 정의 및 전처리 |
| II. 이론적 배경 | IV. 예측모형 개발 및 규칙 도출 |
| 2.1. 항공기 가동률의 중요성 | 4.1. 기상 조건 연계 비행 예측모형 개발 |
| 2.2. 기상 조건과 비행의 연관성 | 4.2. 기상 조건에 따른 비행규칙 도출 |
| 2.3. 데이터마이닝 기법 | 4.3. 비행 규칙 해석 |
| III. 연구 방법 | V. 결론 |
| 3.1. 연구 프레임워크 | 참고문헌 |
| 3.2. 데이터 수집 방법 | <Abstract> |

I. 서론

군용 항공기의 가동률은 무기체계의 작전과 운용을 위한 필수적 요소이며, 항공기가 불가동 상황에 직면할 시 계획되어 있는 임무 수행에 악영향을 끼치게 된다. 항공기 가동 가능 상황은 지속적인 작전을 수행할 수 있도록 전투가 준비된 상태를 의미하며, 최대한 임무에 대한 비행 작전을 수행할 수 있도록 만반의 준비 태세를 갖추어야 한다는 것과 일맥상통한다. 따라서 군용 항공기를 최적의 전투 준비상태로 유

지하는 것만이 대한민국 영공을 지킬 수 있는 절대적인 조건이기도 하다.

국내 군수체계종합업체인 K사가 생산한 군 운영 항공기는 조종사 양성을 목적으로 하는 훈련기와 전투를 목적으로 하는 전투기로 크게 나뉜다. 훈련을 목적으로 하는 훈련기가 불가동 상태인 경우, 조종사 훈련 기간의 연장은 물론이며 극단적으로는 계획한 규모의 조종사 양성도 불가하여 공군의 전력 유지에 문제를 야기하게 된다. 아울러 전투를 목적으로 하는 항공기의 경우는 불가동 시 자국 영공의 방위 전

* 한국항공우주산업, yocrazy@koreaero.com (주저자)

** 하나카드, youngjoo.moon@gmail.com (공동저자)

*** 경상국립대학교 경영대학 경영정보학과, dyjeong@gnu.ac.kr (교신저자)

투력에 크나큰 차질을 주기 때문에 훈련기의 경우보다도 가동률의 중요성은 더욱더 부각 될 수밖에 없다.

항공기의 불가동을 유발하는 요소는 여러 가지 요인이 있겠지만, 본 연구에서는 군 데이터의 특성인 기밀성과 보안성을 고려하여, 기상정보와 연계된 부분으로 한정하여 기상 상태에 따른 군용 항공기의 비행 가능 상태와의 상관관계와 영향성 분석을 연구하고자 한다.

기상 상태가 항공기 운항에 미치는 영향성을 선행적으로 연구한 내용은 풍속, 시정, 강우량 등 기상 현상이 항공기의 지연 여부 및 지연시간에 미치는 영향성을 분석하였으며(최영은, 양봄이, 2021), 활주로 상의 바람의 방향, 돌풍의 정도가 조종사가 착륙을 시도할 때 비행기의 고도 및 속도를 결정해야 하는 중요 요소(이휘영, 2010)임을 연구하는 등 기상 조건이 비행에 지대한 영향을 주는 매우 중요한 요소임을 선행의 연구 등을 통해 알 수 있다.

이에, 본 연구는 국내 군수체계종합업체에서 개발, 생산하여 전력화한 국산 군용 항공기를 대상으로 기상 조건에 따른 비행의 수행과 취소에 대한 연관성을 분석하고 최적의 비행 예측모형을 개발하고자 한다. 그리고 기상정보와 연계한 비행규칙을 도출하여, 기상 데이터를 고려한 효율적인 비행계획 수립 방안을 제시하고자 한다. 본 연구의 목표를 달성하기 위하여 기상청으로부터 기상정보를 수집하였으며, 과거 두 곳의 군 운영 공항에서 축적된 비행 정보와 각 공항의 기상정보를 토대로 비행예측모형을 만들기 위해 로지스틱스 회귀(Logistics Regression), AdaBoost, KNN, Random Forest, LightGBM 등의 기법을 활용하였다. 그리고 이

들 다섯 가지 알고리즘의 예측 정확도, 정밀도, 재현율, F1 Score, AUC_ROC 점수를 종합하여 가장 성능이 좋은 알고리즘을 최종적으로 선정하여 피처(Feature)의 중요도 및 규칙을 도출하였다. 도출된 기상정보와 연계된 비행규칙에 따라 향후 군용 항공기 운항에 대한 비행계획 수립 시 예측된 기상 상황과 연계한 효율적인 계획 수립이 가능토록 대안을 제시하고자 한다.

II. 이론적 배경

2.1. 항공기 가동률의 중요성

미래 전투 환경에 있어 공중 전투력의 중요성은 날로 증가하고 있으며, 이 공중전력을 구성하는 수많은 무기체계 중 가까운 미래의 공중위협에 대응하기 위한 무기는 전투기라 할 수 있겠다. 따라서 공중전력의 근간이 되는 전투기는 국내 독자 연구개발을 통한 전력 강화 혹은 해외 판매업체를 통한 구매 도입 등 다각적 방법으로 지속적인 전력 증강이 필요하며, 획득한 무기가 최상의 가동률을 유지할 수 있도록 철저한 관리 또한 중요하다.

항공기가 고성능 및 고신뢰도를 충족하여 높은 가동률을 유지할 수 있는 것은 미래형 전투체계에 있어 매우 주요한 목표 중 하나이다. 어떠한 무기체계도 운용 시에 고장은 발생이 되는데, 이러한 고장을 사전에 예측할 수 있다면 신속한 군수품 공급이 가능하여 불가동 상태를 최소화할 수 있고, 전체 수리시간도 단축할 수 있어 가동률을 향상시킬 수 있는 효과적인 개선의 수단이다(Macheret et al., 2005).

아울러, 3차원 공간에서 고난이도 기동을 통해 임무를 수행하므로 아무리 미미한 고장이라 할지라도 한번 고장이 발생하면, 경미하게는 임무포기(abort)부터 크게는 항공기 및 조종사의 손실로 이어지는 치명적이고 막대한 결과를 초래할 수 있기 때문이다. 따라서 항공기의 완벽한 임무 수행 여건 보장을 위해 이륙 전 100% 무고장 상태를 보장해야만 한다(정치영, 이재영, 2010). 군용 항공기의 가동률이 중요한 만큼 민간 항공기 가동 상태 또한 운항사업을 수행하고 있는 민간기업의 입장에서 매우 중요한 요소이다. 여객 수요 증가에 따른 수익성을 극대화하기 위해 항공기 가동률 증대의 중요성을 제시하였다(이미숙, 김병중, 2013).

2.2. 기상 조건과 비행의 연관성

착륙 허가국에 의해 지정된 최소 착륙에 필요한 기상 상황을 기준으로 하여 조종사는 착륙에 필요한 최소한의 상황을 각 나라 또는 각 공항에 따라 현 기상을 비교하여 신중히 착륙 여부를 결정해야 한다. 활주로 상의 바람의 방향, 돌풍의 정도는 조종사가 착륙을 시도할 때 비행기의 고도 및 속도를 결정하는 중요 요소이다(이휘영, 2010). 민항기 개발 비행시험을 위한 최적의 시험장소 선정 시 시계(視界), 기상 조건 등을 고려해야 한다. 특히, 양호한 기상 조건과 기상 요인(측풍, 운고, 시정, 바람 등)에 의한 비행 제한 일수가 최소인 비행장을 선정하여야 한다(김찬조 등, 2010). 항공기는 기상 조건과 운항 조건에 따라 엔진 추력, 항공기 이륙 상승률, 이착륙 속도와 거리 증감 등이 발생하고 이로 인하여 항공기 소음도의 변화가 발

생된다. 이준호(2011)는 항공기 소음에 영향을 미치는 외부요인, 기상 조건과 항공기 운항 조건을 선정하여 항공기 소음도 증감을 비교하였으며, 항공기 소음은 기상 조건과 기령과 같은 외부 조건보다는 항공기 이·착륙 중량에 따른 추력 등 항공기 자체 조건에 따라 발생 되는 영향이 더 큰 것으로 확인되었다.

공항에서 발생하는 악천후로 인해 비행 지연 및 항공기 부족이 불가피한 경우 항공사에 막대한 경제적 비용 피해가 발생할 수 있기 때문에 Mao et al.(2015)은 기상 조건에 대비한 항공편 취소 결정 방법론을 제안하였다. 최영은과 양봄이(2021)는 2019년 1년간 제주국제공항의 항공기 운항일지(Tower-log) 자료와 항공 기상청의 제주국제공항 공항기후통계 자료를 활용하여 공항의 운영 특성(일일 운항 횟수, 시간당 운항 횟수 등)과 각종 기상 현상(풍속, 시정, 강수량 등), 항공기 운항 특성(운항노선, 항공기 탑승률 등)이 항공기의 지연 여부 및 지연시간에 미치는 영향을 분석하였다. 그 결과, 일일 운항 횟수, 시간당 운항 횟수, 운항 시간대, 강수량, 풍속, 운항노선, 항공기 탑승률과 같은 다양한 변수가 항공기 지연에 통계적으로 유의한 영향을 미치는 것으로 나타났으며, 변수의 상대적 중요성은 일일 운항 횟수, 풍속, 시정, 항공기 탑승률의 순으로 확인되었다.

2.3. 데이터마이닝 기법

데이터 마이닝(data mining) 기법은 대규모 데이터 세트에서 흥미롭고 예상치 못한 또는 가치 있는 구조를 발견하는 것입니다(Hand, 2007). 데이터 마이닝 기법은 탐색적 자료 분석,

다변량 분석, 시계열 분석, 선형회귀분석, OLAP(On-Line Analytic Processing), SOM (Self-organizing Map), 의사결정나무 분석, 신경망(Neural Network), 머신러닝(Machine Learning) 등의 다양한 기술적인 방법론이 사용된다(최종후, 서두성, 1999). 단, 자료에 의존하여 현상을 해석하고 개선하려고 하기 때문에 자료가 현실적으로 충분히 반영되지 못한 상태에서 정보를 추출한 경우 부적합한 모형을 개발할 수 있는 오류를 범할 수는 있다(이재규 등, 2005).

데이터마이닝 기법은 빅데이터로부터 의미 있는 규칙과 통찰력을 추출하기 위한 컴퓨터 과학 분야이며, 패턴을 결정하거나 발견하기 위해 데이터를 분석하는데 사용된다(Written and Frank, 2005). 의미 있는 규칙은 의사결정나무(decision tree) 형태로 가시화시킬 수 있다. 이러한 데이터마이닝 기법은 제조, 유통, 금융, 통신, 의료 등 다양한 분야에 적용되고 있으며, 분류(classification), 군집(clustering), 연관관계(association), 예측(prediction) 등 다양한 기법을 활용한 분석이 가능하다. 데이터마이닝 기법 중에서 가장 직관적이면서도 분석 결과를 쉽게 해석할 수 있는 기법은 의사결정나무 분석기법이다. 최근에는 LightGBM 알고리즘이 Random Forest보다도 성능 측면에서 우수한 점을 보이면서 두각을 나타내고 있다.

기업은 방대한 양의 데이터가 풍부하나 데이터로 추출된 정보의 활용은 미약하다. Mustafa(2020)는 기업이나 기관이 보유한 방대한 데이터는 귀중한 자원이기 때문에 의사결정에 효과적으로 활용하기 위해서는 통계적 접근 방식으로 접근할 필요가 있으며, 다량의 샘플인

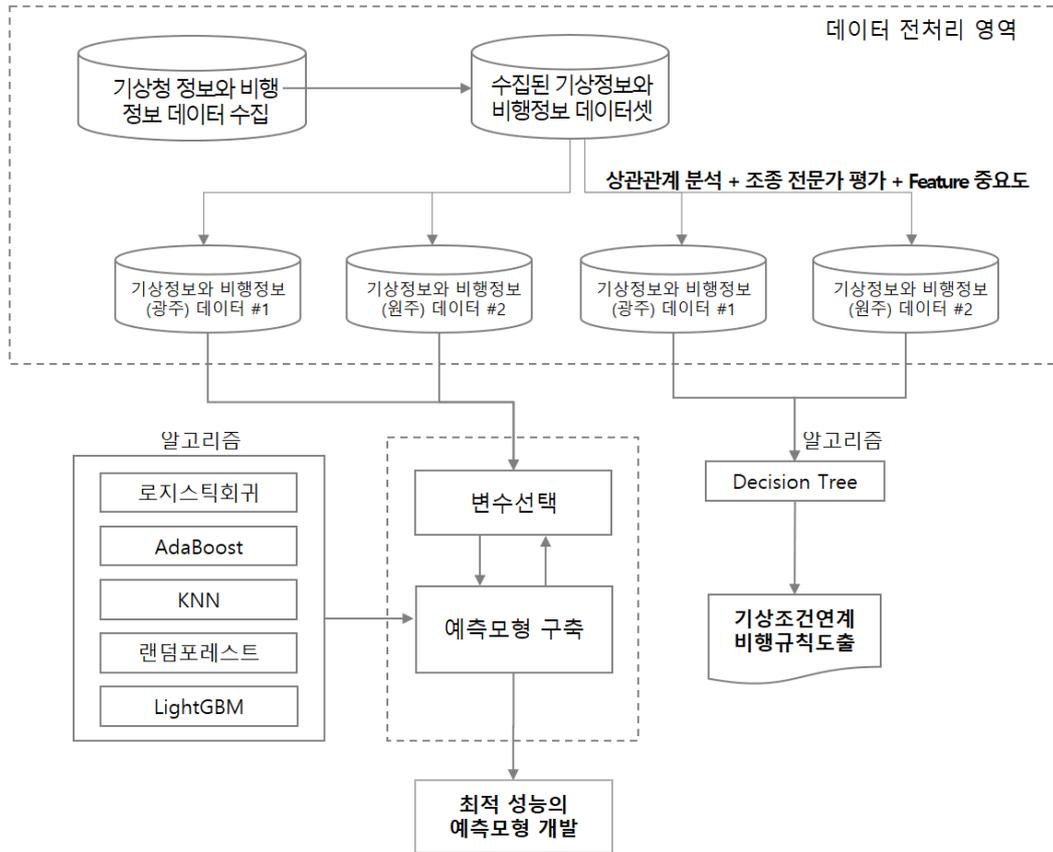
경우는 WEKA 툴을 활용할 것을 권장하였다. 최근 WEKA를 이용하여 다양한 방법으로 빅데이터를 분석한 연구들이 쏟아져 나오고 있다. WEKA와 의사결정나무 분석기법을 이용하여 빅데이터 기반의 예측모형을 개발한 연구로는 최근호 등(2015), 김량형 등(2016), 이동욱(2017), 김덕현 등(2019)이다.

III. 연구 방법

3.1. 연구 프레임워크

본 연구는 대한민국 공군에서 운영하여 축적된 군용 항공기의 비행 실적자료와 기상청에서 제공하는 기상정보를 기반으로 데이터 마이닝 기법을 활용한 해석적 연구이다. 이를 위해 본 연구에서는 국내에서 독자적으로 개발한 군용 항공기의 비행 실적자료와 군용 항공기가 운영되고 있는 인근 지역의 기상청 포털에서 획득한 기상정보를 분석하여 비행의 수행 가능 조건에 대한 최적의 예측모형을 개발하고, 기상 상황에 따른 비행의 규칙을 도출하고자 한다.

본 연구를 수행하기 위한 절차와 연구 프레임워크 제시하면 <그림 1>과 같다. 먼저, 실제로 비행이 이루어지고 있는 2개의 공항(광주, 원주)에서 약 3년간의 군용 항공기의 비행 실적에 대한 방대한 자료를 수집하여 연구 목적에 맞게 전처리를 하였으며, 최종 종속변수인 비행 훈련 스케줄별 비행여부를 정리하였다. 또한, 비행장 인근의 기상정보를 파악하기 위하여 기상청에서 제공하는 공공데이터를 추출하여 데이터 셋을 구성하였다. 또한 분석에 적합한 변



<그림 1> 연구과정 프레임워크

수를 선정하기 위하여 경험이 풍부한 비행 조종 전문가들의 의견을 수렴하였다.

최종적으로 분석에 사용될 변수들을 선택하고 비행훈련 스케줄과 비행장 주변의 기상정보를 이용하여 예측모형을 만들었다. 예측모형 구축을 위해 활용한 알고리즘으로는 로지스틱스 회귀(Logistics Regression), AdaBoost, KNN, Random Forest, LightGBM 등을 활용하였다. 그리고 이들 다섯 가지 알고리즘의 예측 정확도, 정밀도, 재현율, F1 Score, AUC_ROC 점수를 종합하여 가장 성능이 좋은 알고리즘을 최종적으로 선정하여 피쳐(Feature)의 중요도 및

규칙을 도출하였다.

3.2. 데이터 수집 방법

기상 상황 정보와 연계한 비행 가능 예측률을 분류해 보기 위해 2013년 1월부터 2015년까지 총 36개월간 군용 항공기 총 44,558회의 비행 실적 데이터와 기상청 포털에서 제공하는 데이터를 추출하여 본 연구에 활용하였다. 2016년 중반 이후의 정보는 비행계획과 취소에 대한 군 운영 정보의 보안 관계로 인하여 확보에 어려움이 있었다. 비행훈련 정보는 국내 군

수체계종합업체인 K사가 제작하여 납품한 군용 항공기 중 훈련기와 공격기를 중심으로 수집되어졌다. 비행훈련 데이터는 광주, 횡성(원주) 공항에서 이루어진 비행 이력 정보이다. 기상정보는 기상청에서 제공하는 공공데이터를 활용하였다. 광주는 기상청에서 제공하는 기상정보를 수집하였으나, 횡성 지역은 제공하지 않아 근접 지역인 원주를 대상으로 기상정보를 수집하였다.

3.3. 연구변수의 정의 및 전처리

기상 상황 정보와 연계하여 최적의 비행 가능성 판단을 위한 예측모형 분석을 위해, 데이터 수집 단계에서는 총 15개의 변수가 고려되었으나 이 중에서 선별하여 최종 분석에는 총 10개의 독립변수와 1개의 종속변수를 선정하였다(<표 1> 참조).

기상청에서 제공하는 강수량에는 비, 눈, 우박 등을 포함하고 있는 값이기 때문에 적설량

<표 1> 수집 데이터

구분	변수명	변수 설명	전처리 후 사용변수	변수 번호
독립변수	기온(°C)	공기의 온도이며 일반적으로 지상 1.5m 높이의 대기온도	○	F1
독립변수	강수량(mm)	강우량과 강설량 등을 통틀어서 지표면에 떨어진 강수의 양(비, 눈, 우박 등 포함)	○	F2
독립변수	풍속(m/s)	바람의 속도	N/A	-
독립변수	풍향	바람이 불어오는 방향이며 16방위로 관측	N/A	-
독립변수	풍향*풍속	풍향을 8방위로 구분 후 풍속을 곱한 값 예) 서풍: F4_W, 북서풍: F4_NW	○	F3
독립변수	습도(%)	공기 중에 포함되어 있는 수증기의 양 또는 비율을 나타내는 단위(통상 상대습도)	○	F4
독립변수	증기압(hpa)	고체 또는 액체에서 증발하는 압력으로 증기나 고체나 액체와 동적 평형 상태에 있을 때의 포화증기압	○	F5
독립변수	현지기압(hpa)	일정 지역에서 관측된 기압 (본 연구에서는 광주, 원주)	○	F6
제거변수	적설(cm)	실제로 지면에 쌓여 있는 눈의 깊이를 말하며, 관측일(보통 00시를 기준으로 시작)	N/A	-
독립변수	전운량(10분위)	하늘을 덮고 있는 구름의 양의 비율을 말하며 단위는 10할로 구분	○	F7
제거변수	중하층운량(10분위)	중층과 하층에 있는 구름의 분포량	N/A	-
독립변수	최저운고(100m)	지(수)면으로부터 구름의 높이로 이륙보다는 착륙시 운고가 낮으면 안전에 위협	○	F8
독립변수	이슬점온도(°C)	이슬점 또는 노점(露點)은 공기가 포화되어 수증기가 응결할 때의 온도 혹은 불포화 상태의 공기가 냉각될 때 포화되어 응결이 시작되는 온도	○	F9
제거변수	이륙시각	비행 이륙 시각	N/A	-
독립변수	시정(10m)	대기의 혼탁한 정도를 나타내는 것으로 정상적인 육안으로 멀리 떨어진 물표를 인식할 수 있는 최대거리를 말함	○	F10
종속변수	비행	비행 여부 (Go/NoGo)	○	F11

에 대한 변수는 중복성으로 제거하였으며, 중하층 운량은 전운량 및 최저운고에 대한 독립변수가 고려되었기 때문에 분석 대상에서 제거하였다. 이륙 시각은 단지 비행 시작의 개념으로 날짜 형태의 데이터이기 때문에, 시계열의 의미로 분석 시 결과의 오류를 유발할 수 있기에 독립변수에서 제거하였으며, 군 운영 특성상 주말은 비행하지 않아 전 기간 중 주말의 기상 조건 데이터는 제거하여 전처리하였다. 정제된 데이터는 기상 시간대별 기상청 정보와 비행 수행과 취소 정보를 시계열로 연동 처리하였고, 종속변수인 비행 여부의 데이터는 바이너리 변수로 변경하여 전처리하였다. 풍속과 풍향 변수는 비행 시 개별적인 영향보다는 상호 복합적인 작용을 하기 때문에, one-hot-encoding 후 풍속을 곱한 벡터값의 유사한 형태로 처리하여 풍향*풍속의 변수로 적용하였으며, 독립변수값은 Robust Scalar를 적용하여 데이터 표준화를 진행하였다.

IV. 예측모형 개발 및 규칙 도출

4.1. 기상정보 연계 비행 예측모형 개발

본 연구에서는 기상정보 연계 비행 예측모형

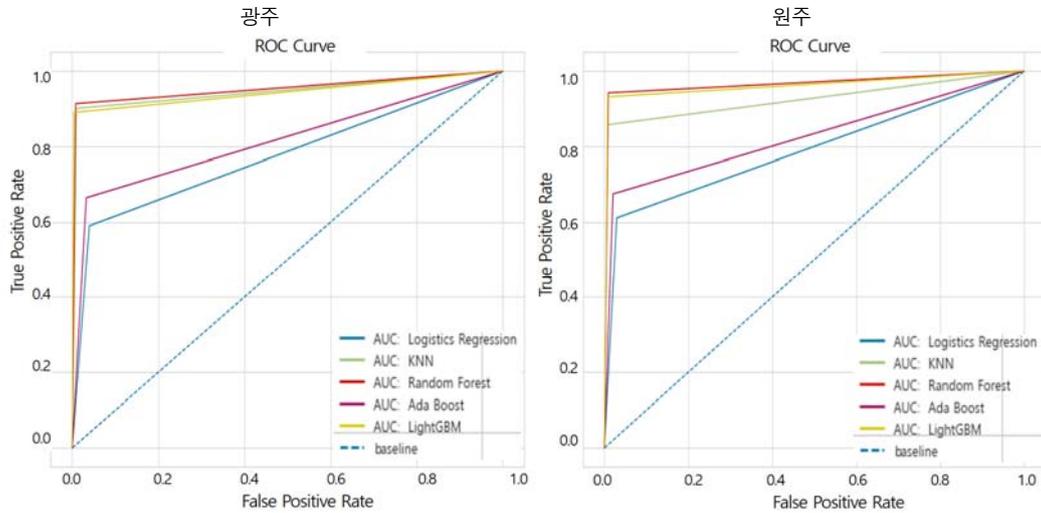
개발을 위해 로지스틱스 회귀(Logistics Regression), AdaBoost, KNN, Random Forest, LightGBM 등의 알고리즘을 활용하였다. 그리고 이들 다섯 가지 알고리즘의 예측 정확도, 정밀도, 재현율, F1 Score, AUC_ROC 점수를 종합하여 가장 성능이 우수한 알고리즘을 최종적으로 선정하여 Feature의 중요도 및 규칙을 도출하였다. 본 연구에서 예측한 모형의 성능 결과는 <표 2>와 같다.

성능 예측 결과 Random Forest와 LightGBM 모형이 가장 성능이 우수했으나, 항공기 비행 취소 예측 모형은 2중 오류를 최소화하여 조종사에게 안전한 비행의 기상 조건을 보장해줘야 하기에, 정밀도 보다는 재현율에 무게중심을 두고 우선순위를 평가하였다. 광주와 원주 모두 LightGBM 알고리즘이 높은 성능을 보여주어 비행 취소 예측을 위한 최적의 분석 모형으로 선정하였다.

<표 2>의 AUC 결과와 <그림 2>의 ROC Curve를 살펴보면, KNN, Random Forest, LightGBM 알고리즘의 성능이 90% 이상으로 성능의 우수함을 확인할 수 있기 때문에, 최적의 분석 모형은 재현율을 고려해 볼 때 LightGBM, KNN, Random Forest 순으로 우수하였다.

<표 2> 예측모형 예측률 성능 결과 비교

모델	데이터 #1 (광주)					데이터 #2 (원주)				
	정확도	정밀도	재현율	F1	AUC	정확도	정밀도	재현율	F1	AUC
LR	89.11	59.02	77.44	66.98	77.53	91.64	61.11	78.57	68.75	79.08
Ada Boost	91.08	66.44	82.49	73.60	81.60	93.35	67.46	85.28	75.33	82.70
KNN	97.36	90.07	95.57	92.74	94.55	97.17	85.71	95.01	90.13	92.46
RF	97.75	91.32	96.47	93.83	95.27	98.33	94.18	94.68	94.43	96.62
LGBM	97.67	88.98	98.43	93.47	94.33	98.29	93.12	95.39	94.24	96.16



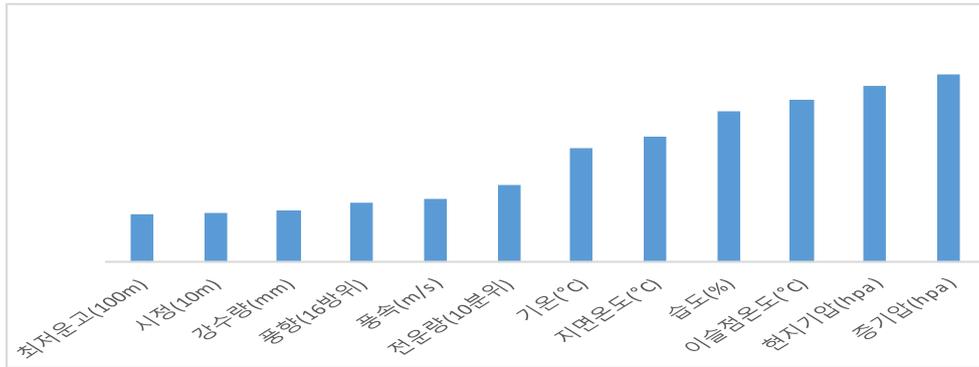
<그림 2> 공항/예측별 ROC Curve

4.2. 기상 조건에 따른 비행규칙 도출

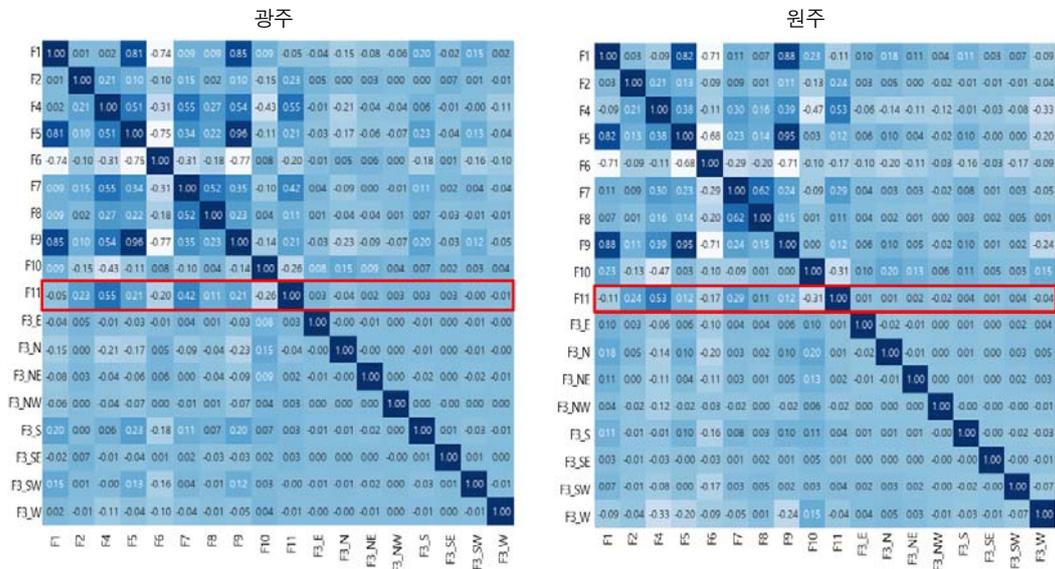
실제 비행기지에서 비행이 가능한 기상 조건은 어떻게 판단되고 활용하는지 조사하기 위해 25년 이상 군용 전투기 조종 전문가들의 의견을 수렴하였다. 정성적인 관점에서 조종 전문가들의 의견은 통상 비행 여부 결정은 기상 제한 조건의 상태로 판단하며, 기상 제한 요소는 운고, 시정, RCR(runway condition reading, 활주로 상태), 측풍 등이 비행에 제한을 주는 요소라고 평가한다. 특히 운고와 시정 조건 상태에 따라 시계비행 수행의 가능 여부를 판단하며, 악 기상 조건에서는 계기비행의 수행은 불가 판정이 될 수 있다. 강수량 자체에 대한 비행 제한은 없으나 비가 많이 오면 시정이 제한될 수 있고, 활주로에 물이 고이게 되면 활주로 상태 저하로 슬립(slip) 현상이 발생되어 비행이 제한될 수 있다. 습도와 지면 온도 조건은 안개와 관련이 있으므로 시정을 예측하는데 활용될 수 있

다. 앞서 기술한 기상 조건에 대한 정성적인 비행 제한 요소 등을 고려하여, 기상 독립변수 중 비행이 불가한 상황으로 영향을 많이 미치는 변수를 1점에서 가장 낮은 영향을 미치는 변수는 12점으로 부여하였다. 조종 전문가들이 부여한 점수의 평균을 적용하여 가장 낮은 점수가 비행 제한 조건의 영향을 가장 크게 주는 변수로 고려하는 방식으로 정량적 평가를 수행하였으며, 결과는 <그림 3>과 같다. 조종 전문가들은 최저운고를 가장 중요한 변수로 보았으며, 시정, 강수량, 풍향에서 증기압의 순으로 비행을 결정하는 중요 요소로 판단하였다.

파이썬의 히트맵(heatmap)을 활용하여 각 변수간 상관관계를 분석한 결과는 <그림 4>와 같다. 광주 공항의 경우는 습도, 전운량, 시정, 강수량, 증기압의 순으로 높은 상관관계를 보여주었으며, 원주 공항은 습도, 시정, 전운량, 강수량, 현지기압 순으로 높은 상관관계로 확인되었다.



<그림 3> 변수 중요도 우선순위 조종 전문가 평가 결과



<그림 4> 변수 중요도 우선순위 평가 히트맵 결과

아울러, 공항별 변수의 중요도는 <그림 5>와 같으며, 광주 공항의 경우 가장 중요한 변수로는 현지기압, 기온, 습도, 증기압, 시정 순이었고, 원주 공항의 경우 가장 중요한 변수로는 현지기압, 기온, 습도, 시정, 증기압 순이었다.

주요 기상 조건에 따른 비행규칙을 도출하기 위해 조종 전문가들의 중요도 평가, 파이썬의

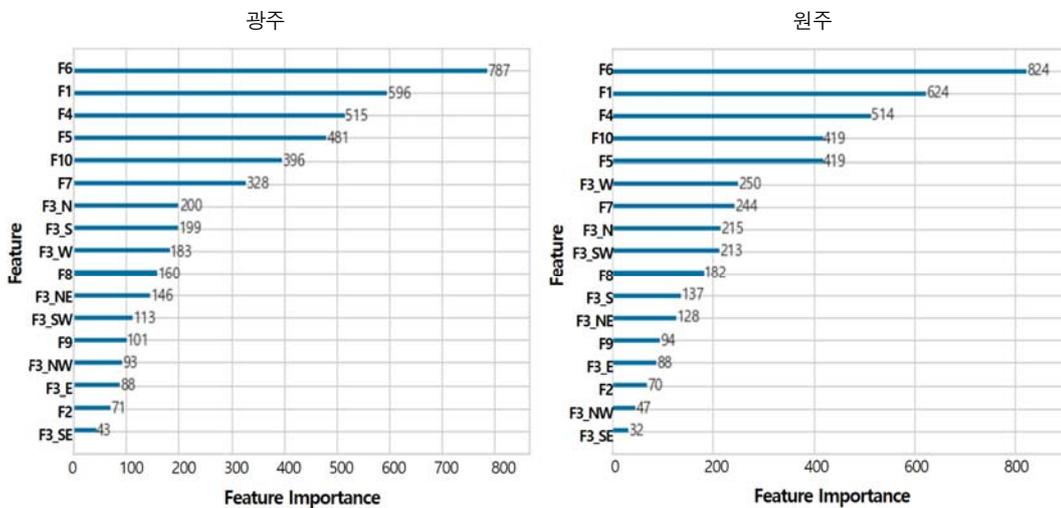
히트맵을 통한 상관관계 분석 결과, 피쳐(feature) 중요도 평가 결과를 종합적으로 고려하여 비행 불가 조건에 가장 큰 영향을 주는 5대 주요 변수를 선별하였다. 조종 전문가가 평가한 중요도가 높은 5대 변수는 최저운고, 시정, 강수량, 풍향/풍속, 전운량이었으며, 히트맵은 습도, 시정, 전운량, 강수, 현지 기압을 중요

변수로, 피처(feature) 중요도 평가에서는 현지 기압, 기온, 습도, 증기압 및 시정의 순으로 선별되었다. 각 중요도 평가 결과에 2개 이상 중복이 되는 영향성이 높은 변수를 선별하였으며, 비행 규칙을 산출하기 위한 최종 변수로 습도, 시정, 전운량, 강수량, 현지기압 변수가 선정되었다.

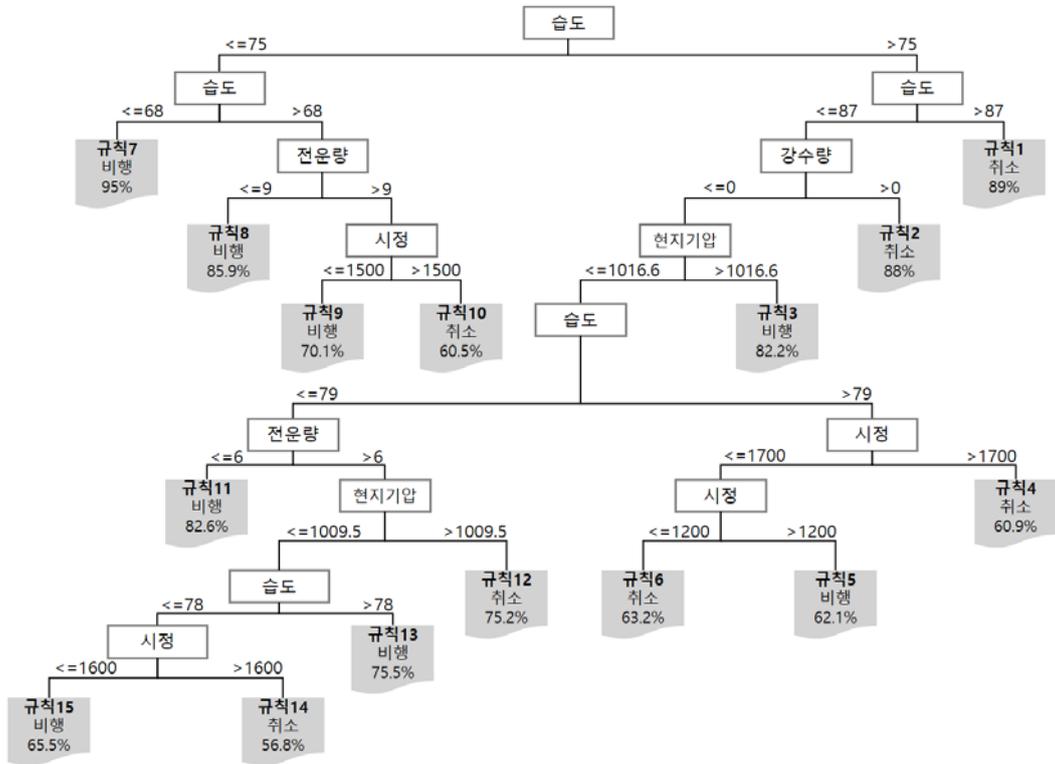
통상 인공지능경망과 같은 알고리즘의 경우는 블랙박스의 형태로 해석 부분이 계산되어 규칙 간의 관계를 파악하기에는 무리가 있다(김원중 등, 2018). 기상 조건에 따른 비행규칙을 도출하기 위해서 비행 예측모형 개발 시 가장 우수한 성능을 보여준 LightGBM tree 알고리즘을 활용을 시도하였으나, 비행 취소 확률이 50% 확률 수준으로 밖에 성능이 확인되지 않아, 비교적 해석이 용이한 의사결정나무(decision

tree) 알고리즘을 활용하였다. 의사결정나무 알고리즘을 통해 도출된 결과는 나무 구조로 표현되기 때문에 가장 설명력 있는 변수가 최초로 분리되는 특징을 가지고 있어 중요한 변수와 함께 비행 취소에 영향을 주는 변수들의 관계를 파악하기 용이하다.

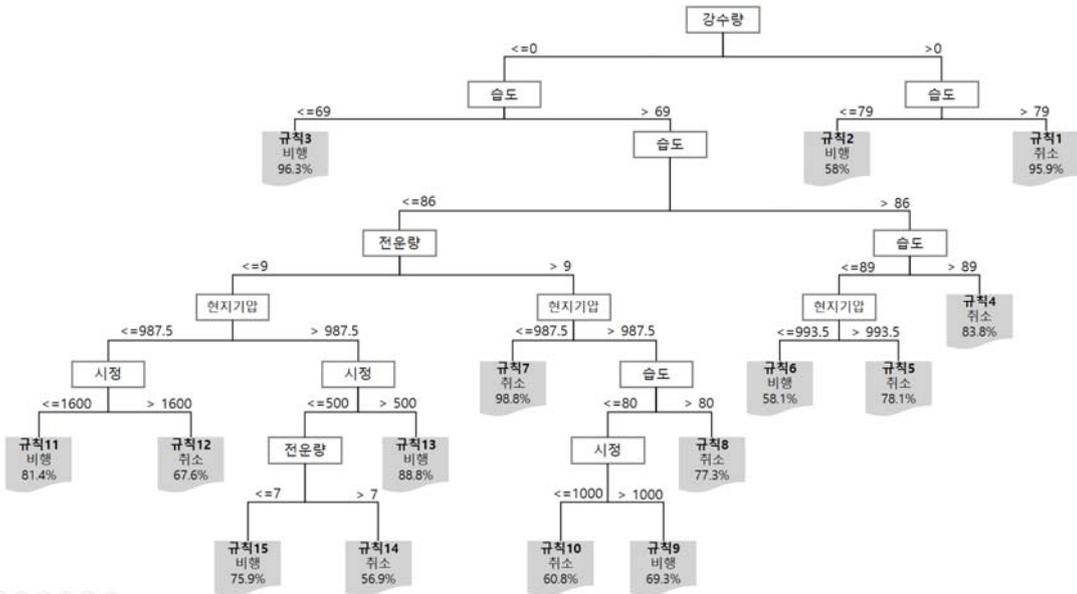
아울러 지역별 기상 조건도 상이할 뿐만 아니라, 공항의 활주로 상황과 항공기 운항 대수의 차이도 있다는 점 등을 고려하여, 최적 예측 모형 구축과 동일하게 데이터셋을 통합하지 않고, 각 공항별 기상 조건에 따른 비행 취소 규칙을 도출하였다. 광주 는 의사결정나무 모형의 적중률은 90%, 원주는 92%의 적중률의 분석 결과, 훈련데이터와 테스트 데이터 비율은 7:3으로 동일하게 진행하였으며, 그 결과는 <그림 6> 및 <그림 7>과 같다.



<그림 5> 공항별 Feature 중요도



<그림 6> 광주 공항 비행규칙 도출 결과



<그림 7> 원주 공항 비행규칙 도출 결과

4.3. 비행 규칙 해석

원주 공항의 비행 환경은 군용 항공기와 민간 항공기가 함께 운용되고 있는 공용공항이며, 항공기 운행은 광주 공항 내 활주로 방향이 남북방향(4R/22L, 4L/22R)으로 위치하고, 대부분의 운행은 남쪽에서 북쪽 방향으로 운영하고 있다(김연풍, 2010). 광주 공항의 위치는 북위 35도, 동경 126도이다. <표 3>은 광주 공항의 기상 조건에 따른 비행규칙에 관한 세부 내용을 설명하고 있다.

비행할 수 있는 조건을 도출하기 위해 비행 취소가 일어나는 기상 조건에 대해 역 조건을 만족하면 비행을 만족하는 조건이므로 비행 취소 규칙에 대해 중점적으로 분석하였다. 광주 공항의 비행 규칙을 살펴보면, 규칙 1은 습도가

87 이상, 즉 우리나라의 여름 습도가 70~85%인 점을 고려할 때, 습도가 매우 높으면 온도가 높을 확률도 있기 때문에 비행을 취소할 확률은 89.0%로 가장 높은 비행 취소 확률임을 알 수 있다. 규칙 2는 습도가 75에서 87사이에서 강수량은 0 이상으로 비가 오게 되면 비행 취소 확률이 88.0%로 높은 수준이다.

규칙 4는 습도가 75에서 87사이에서 강수량은 0 이하지만 현지기압이 1,016.6보다 작을 시 비행 취소 확률이 60.9%이다. 규칙 6은 규칙 4와 비슷한 조건에서 시정이 1,200 이하이면 취소 확률이 63.2%이다. 규칙 10은 습도가 68이상 75이하로 우리나라 연중 평균 습도의 수준이나, 시정이 1,500이상이나 전운량이 9분위 이하이면 취소 확률이 60.5%이다. 규칙 12는 한 여름의 습도 조건과 전운량이 6분위 이상일 때

<표 3> 광주 공항 비행 취소 규칙

규칙	습도	강수량	현지기압	시정	전운량	비행여부	취소확률
1	>87					취소	89.0%
2	75<h<=87	>0				취소	88.0%
3	75<h<=87	<=0	>1,016.6			비행	82.2%
4	79<h<=87	<=0	<=1,016.6	>1,700		취소	60.9%
5	79<h<=87	<=0	<=1,016.6	1,200<v<=1,700		비행	62.1%
6	79<h<=87	<=0	<=1,016.6	>1,200		취소	63.2%
7	<=68					비행	95.0%
8	68<h<=75				<=9	비행	85.9%
9	68<h<=75			<=1,500	<=9	비행	70.1%
10	68<h<=75			>1,500	<=9	취소	60.5%
11	75<h<=79	<=0	<=1,016.6	>1,700	<=6	비행	82.6%
12	75<h<=79	<=0	1,000.9<p<=1,016.6		>6	취소	75.2%
13	78<h<=79	<=0	<=1,009.5		>6	비행	75.5%
14	75<h<=78	<=0	<=1,009.5	>1,600	>6	취소	56.8%
15	75<h<=78	<=0	<=1,009.5	<=1,600	>6	비행	65.5%

현지기압이 1,000.9에서 1,016.6일 때 비행 취소 확률이 75.2%이다. 광주 연평균 현지기압이 1.008임을 고려할 때 현지기압의 영향보다는 습도와 전운량의 영향으로 판단된다. 규칙 14는 한 여름의 습도 조건, 시정은 1,600으로 비교적 양호하나 현지기압이 1,009.5이하이고 전운량이 6분위 이상을 보일 때 비행 취소 확률은 56.8%이다.

광주 공항에서 비행이 취소될 확률의 규칙을 종합하면, 습도가 87% 이상으로 한 여름 정도로 습하거나 비가 내릴 경우 비행 취소가 될 확률이 90%에 근접할 정도로 매우 높았으며, 통상 전운량은 7~8분위 수준일 때 시계비행의 제약으로 비행 취소 확률이 발생할 수 있는데, 광주 공항은 훈련을 통한 전투기 조종사를 양성하는 특성으로 전운량이 6분위 이상 흐린 날씨

도 비행 취소될 확률이 75% 수준으로 취소됨을 알 수 있었다.

원주 공항의 비행 환경은 군용 항공기와 민간 항공기가 함께 운용되고 있는 공용공항이며, 위치는 북위 37도, 동경 127.9도이며, 항공기 운행은 북동남서 방향(03/21)으로 활주로가 배치되어 있다. <표 4>는 원주 공항의 기상 조건에 따른 비행규칙에 관한 세부 내용을 설명하고 있다.

원주 공항의 비행 취소 규칙은, 규칙 1을 보면 한여름의 습도가 유지되면서 비가 내리는 경우 비행 취소 확률은 95.9%로 절대적으로 가장 높은 비행 취소 확률임을 알 수 있었다. 규칙 4는 강수량은 적지만 습도가 89% 이상이면 비행 취소 확률이 83.8%로 매우 높은 비행 취소율을 나타냄을 알 수 있다. 규칙 5는 강수량은

<표 4> 원주 공항 비행 규칙

규칙	강수량	습도	현지기압	시정	전운량	비행여부	취소확률
1	>0	>79				취소	95.9%
2	>0	<=79				비행	58.0%
3	<=0	<=69				비행	96.3%
4	<=0	>89				취소	83.8%
5	<=0	86<h<=89	>993.5			취소	78.1%
6	<=0	86<h<=89	<=993.5			비행	58.1%
7	<=0	69<h<=86	<=987.5		>9	취소	98.8%
8	<=0	80<h<=86	>987.5		>9	취소	77.3%
9	<=0	<=80	>987.5	>1,000	>9	비행	69.3%
10	<=0	<=80	>987.5	<=1,000	>9	취소	60.8%
11	<=0	69<h<=86	<=987.5	<=1,600	<=9	비행	81.4%
12	<=0	69<h<=86	<=987.5	>1,600	<=9	취소	67.6%
13	<=0	69<h<=86	>987.5	>500		비행	88.8%
14	<=0	69<h<=86	>987.5	<=500	>7	취소	56.9%
15	<=0	69<h<=86	>987.5	<=500	<=7	비행	75.9%

적지만 습도가 86%에서 89%의 한여름 날씨의 습도로 매우 높아 비행 취소 확률은 78.1%로 높은 수치임을 알 수 있다. 규칙 7은 강수량은 거의 없지만 습도가 거의 한여름 시기의 습도 조건으로 전운량이 9분위 이상이면 비행 취소 확률이 98.8%로 최고의 비행 취소 조건임을 알 수 있었다. 규칙 10은 강수량이 적고 습도도 여름 조건의 동등 수준으로 유지되고 있고, 다만, 현지기압이 987.5보다는 크지만, 시정 조건이 1,000미터 미만이고, 전운량이 9분위 이상으로 비행 취소 확률이 60.8%임을 보았을 때, 원주 공항은 시정의 문제보다는 습도와 전운량에 비행 취소 확률 영향이 있음으로 판단할 수 있겠다.

원주 공항에서 비행이 취소될 확률의 규칙을 종합하면, 비가 내리거나, 혹은 강수량은 적으나 습도가 80% 이상, 전운량이 9분위 이상으로 매우 습한 날씨이고 구름량이 전체적으로 공역을 덮고 있는 환경일 때 비행이 취소되는 규칙을 발견할 수 있었다. 광주 공항 분석 결과에서도 언급은 하였지만 원주 공항의 경우에는 학생 조종사의 기술이 아닌 전문적인 전투기 조종사의 기술을 보유하고 있어 전운량의 경우는 9분위 이상의 경우에서 비행의 수행과 취소가 분류됨을 알 수 있었다. 반면, 강수량이 적고 습도도 70% 미만으로 유지될 때 비행 확률이 96.3%로 가장 최적의 조건임을 알 수 있었다.

광주 및 원주 공항의 비행 취소 규칙을 종합하면, 강수량이 미세하게라도 발생하는 경우 시계비행에 어려움을 주기 때문에 계기비행이 미흡한 학생 조종사들에게는 좋지 않은 상황이고, 활주로 상에 수막이 형성되어 이착륙시 안전에 위협 요소가 될 수 있어 비행 취소에 영향이 큰

변수로 작용하였다. 습도는 80% 수준일 경우 비행을 하기 위한 최적의 기상 조건에 제약이 발생할 확률이 높다는 것을 알 수 있었다. 전운량이 9분위 이상으로 날씨가 매우 흐린 상황이면 구름, 안개 등 조종사 시정에 악영향과 관련될 수 있는 운영환경으로 비행에 지장을 초래할 수 있는 기상 조건이 될 수 있음을 알 수 있었고, 특히 학생 조종사들을 양성하는 광주 공항은 전운량이 6분위 이상만 되어도 비행 취소 발생 확률이 있음을 알 수 있었다. 습도가 높고 전운량이 매우 높다는 의미는 최저온고 역시 낮을 수 있는 확률로 비행 착륙 시 시정을 확보할 수 없어 조종이 미숙한 학생 조종사의 경우는 비행 안전에 위협적인 요소로 작용될 수 있다. 광주와 원주 공항의 경우는 강수량, 습도, 전운량에 비해 시정은 비행 취소에 비교적 낮은 영향성을 보임을 알 수 있었다.

V. 결론

본 논문은 군용 항공기를 대상으로 기상 조건에 따른 비행의 규칙을 도출하였다는 것에는 의미가 있다. 기상 조건에 따른 비행 가능성을 예측할 수 있는 최적의 분류모형을 도출한 점과 decision tree 알고리즘을 활용하여 기상 조건에 따른 비행규칙을 연구하였다는 것에 의의가 있겠다. 특히 선행 논문에서는 기상 조건을 고려해야 된다는 점을 강조하는 것으로 언급이 되었지만, 본 논문에서는 습도, 시정, 전운량, 강수량 및 현지기압 등 세부적인 기상 조건과의 상관관계로 비행의 취소가 발생할 수 있다는 결과를 도출한 것은 접근 방법론 측면에서

도 의미가 있다. 또한, 일반적으로 비행 취소 영향을 주는 조건에 조종 전문가 평가를 통해 시정에 대한 영향도 높을 것으로 인지되었지만, 비행 취소 여부 측면으로의 데이터마이닝 접근에서는 기타 변수보다는 비행 취소의 영향에 있어 미미한 것으로 확인되었다는 점도 의미가 있다.

항공기는 양력에 의해서 이착륙이 되는 원리이며, 공기의 밀도는 온도, 압력, 습도에 의해 변화된다. 습도가 높으면 공기밀도가 낮아지고 이는 엔진 성능, 기동성, 착륙 성능이 떨어진다는 결과를 초래하기 때문에 이륙거리도 길어지게 되지만, 항공기의 비행 여부는 기상 조건으로만 결정되는 것은 아니다. 수리를 위해 필요한 군수품의 부족이 비행을 취소할 수 있는 경우도 많기 때문이다. 다만 항공기가 가동 가능한 상태일 때를 기준으로 수립한 비행계획을, 기상 조건과 연계하여 비행의 취소 여부를 확인하였기 때문에 물자의 부족으로 인한 항공기 불가동 상태는 기해소된 결과가 반영되어 해석의 오류를 범한 것은 아니라고 판단된다. 일반적으로 민항기의 경우는 보잉이나 에어버스 등 대형 항공사가 제작한 항공기를 운영하고 있어, 물자 공급망(supply chain)이 군용 항공기보다는 융통성이 많아 물자 부족으로 인한 비행의 취소 영향성은 군용 항공기보다는 미미할 것으로 판단된다. 본 연구에서 도출한 비행의 규칙을 고려하여 비행계획을 수립한다면 군용 항공기를 보다 체계적이고 실효성 있게 작전과 임무 수행이 가능할 것으로 기대된다.

지금까지 기상 조건과 관련하여 진행된 유사 연구들은 대부분 민간 항공기 분야에서 연구가 진행되었지만, 본 연구가 추구하고 있는 방향성

과 기존 연구와의 차이점을 정리하면 다음과 같다. 첫째, 대한민국이 독자적으로 개발하여 제작한 군용 항공기의 비행에 영향을 주는 주요한 기상 조건들의 상관관계를 분석하였다. 둘째, 기상청에서 제공하는 기상 포털 데이터와 군용 항공기의 실 운영데이터를 반영하여 분석함으로써 분석 결과의 건전성과 신뢰도가 높다. 셋째, 다차원 분석에 보다 탁월한 데이터 마이닝 기법을 활용하여 기상 요인들 간의 관계 분석을 통한 최고 성능의 예측모형을 개발하였고, 25년 이상 군용 항공기 조종 전문가들의 의견을 반영하여 기상정보와 연계한 비행규칙을 도출하는 연구를 수행하였다. 본 연구의 결과는 군용 항공기 비행계획 수립 시 도움을 줄 것으로 기대한다.

기상의 조건만을 고려하여 기상청의 데이터와 일부 군용 항공기의 비행 여부를 연계하여 분석 결과를 도출하다 보니, 현재 연구된 군용 항공기 기종과 다른 기종에 대해서 일반화하기 위해서는 추가적 연구가 필요하며, 특히 민간 항공기의 비행규칙으로의 적용을 일반화하는데에는 본 연구의 한계점으로 인식된다. 최근 지구 환경 변화로 인해 우리나라 기온이 고온 다습한 열대야성 기후로 변화가 되고 있다는 점을 고려해 본다면, 향후 기상정보를 추가로 확보하여 최신화된 비행규칙 연구가 필요할 것으로 사료된다.

참고문헌

김덕현, 유동희, 정대율, “의사결정나무 기법을 이용한 노인들의 자살생각 예측모형 및

- 의사결정 규칙 개발,” 정보시스템연구, 제28권 제3호, 2019, pp.249-276.
- 김량형, 유동희, 김진우, “데이터마이닝 기법을 이용한 기업부실화 예측 모델 개발과 예측 성능 향상에 관한 연구,” *Information Systems Review*, 제18권 제2호, 2016, pp.173-198.
- 김원중, 최연식, 유동희, “데이터 마이닝을 활용한 한국 프로야구 구단의 승패 예측과 승률 향상을 위한 전략 도출 연구,” *한국스포츠산업경영학회지*, 제23권, 제3호, 2018, pp.88-104.
- 김연풍, “공항주변 소음도 분포특성에 관한 연구,” 박사학위 논문, 조선대학교 대학원 환경생명공학과, 2010.
- 김찬조, 남기은, 우봉길, “민항기 개발 비행시험을 위한 최적 시험장소 연구,” *한국우주공학학회 학술대회 발표논문집*, 2010, pp.1286-1289.
- 이미숙, 김병중, “항공 수요 증가에 따른 항공사 운영 생산성 및 수익성 변화 추정,” *교통연구*, 제20권, 제4호, 2013, pp.55-66.
- 이동욱, “데이터 마이닝을 활용한 자동차 재구매 증진 방안에 관한 연구,” 석사학위 논문, 경상대학교 경영정보학과, 2017.
- 이재규, 권순범, 임규건, *경영정보시스템 원론* (제2판), 법영사, 2005.
- 이준호, “외부요인에 의한 항공기 소음도 증감에 관한 연구,” *항공진흥*, 제57호, 2011, pp.21-40.
- 이휘영, “항공사 수입관리를 위한 효율적인 운항관리 방안에 관한 연구,” *항공진흥*, 제55호, 2010, pp.67-88.
- 정치영, 이재영, “연관성 분석 기법을 활용한 항공기 결함 분석,” *Journal of the Korean Data Analysis Society*, 제12권, 제1호, 2010, pp.261-269.
- 최근호, 서용무, 유동희, “산재근로자들의 고용안정과 건강한 삶을 위한 데이터 마이닝 기반의 규칙 도출 연구,” *직업재활연구*, 제25권, 제3호, 2015, pp.5-24.
- 최영은, 양봄이, “제주국제공항의 항공기 지연 요인 분석,” *대한교통학회지*, 제39권 제2호, 2021, pp.137-148.
- 최중후, 서두성, “데이터마이닝 의사결정나무의 응용”, *통계분석연구*, 제4권, 제1호, 1999, pp.61-83.
- Macheret, Y., Koehn, P. and Sparrow, D., *Improving Reliability and Operational Availability of Military Systems*,” *IEEE Aerospace Conference*, 2005, pp.3948-3957.
- Hand, David J. *Principles of data mining*. Drug Safety, 2007, pp. 621-622
- Mao, X., Yang, Y., Cai, K. and Yang W., “A Decision Support Method for Flight Cancellation in Adverse Weather: An Airport Perspective,” *IEEE Digital Avionics Systems Conference(DASC)*, 2015, pp.1E2-1-1E2-9.
- Mustafa, R. H., “Big Data Analysis using WEKA Machine Learning Program and SPSS Package: A Comparative Study,” Degree of PhD in Statistics, SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY, 2020.

Song, Y. and Lu, Y., "Decision Tee Methods: Applications for Classification and Prediction," Shanghai Arch Psychiatry, Vol. 27, No. 2, 2015, pp. 130-135.

Written, I. H., and Frank, E. *Data Mining: Practice Machine Learning Tools and Technique*, San Francisco: Morgan Kaufmann Publishers, 2005.

유 경 열 (Yu, Kyoung Yul)



경상국립대학교 대학원 경영정보학과 박사과정을 수행 중이다. 현재 한국항공우주산업(주) ICT융합실장으로 재직하고 있으며, 주요 관심분야는 스마트 제조 플랫폼, 빅데이터, 사무자동화(RPA) 등이다.

문 영 주 (Moon, Young Joo)



미국 퍼듀대학교 경영학사를 취득하였고, 성균관대학교 데이터사이언스융합학과 석사과정을 수행 중이다. 현재 하나카드 데이터사이언스팀에 재직하고 있으며, 주요 관심분야는 정형/비정형 데이터 분석, 그래프 마이닝, 추천 시스템 등이다.

정 대 울 (Jeong, Dae Yul)



부산대학교 경영학사, 석사, 박사학위를 취득하였고, 한국정보시스템학회 회장을 역임하였다. 현재 경상국립대학교 경영정보학과 교수로 재직하고 있으며, 주요 관심분야는 시스템분석 및 설계, 데이터마이닝, 의사결정지원시스템 등이다.

<Abstract>

A Study on the Development of Flight Prediction Model and Rules for Military Aircraft Using Data Mining Techniques

Yu, Kyoung Yul · Moon, Young Joo · Jeong, Dae Yul

Purpose

This paper aims to prepare a full operational readiness by establishing an optimal flight plan considering the weather conditions in order to effectively perform the mission and operation of military aircraft. This paper suggests a flight prediction model and rules by analyzing the correlation between flight implementation and cancellation according to weather conditions by using big data collected from historical flight information of military aircraft supplied by Korean manufacturers and meteorological information from the Korea Meteorological Administration. In addition, by deriving flight rules according to weather information, it was possible to discover an efficient flight schedule establishment method in consideration of weather information.

Design/methodology/approach

This study is an analytic study using data mining techniques based on flight historical data of 44,558 flights of military aircraft accumulated by the Republic of Korea Air Force for a total of 36 months from January 2013 to December 2015 and meteorological information provided by the Korea Meteorological Administration. Four steps were taken to develop optimal flight prediction models and to derive rules for flight implementation and cancellation. First, a total of 10 independent variables and one dependent variable were used to develop the optimal model for flight implementation according to weather condition. Second, optimal flight prediction models were derived using algorithms such as logistics regression, Adaboost, KNN, Random forest and LightGBM, which are data mining techniques. Third, we collected the opinions of military aircraft pilots who have more than 25 years experience and evaluated importance level about independent variables using Python heatmap to develop flight implementation and cancellation rules according to weather conditions. Finally, the decision tree model was constructed, and the flight rules were derived to see how the weather conditions at each airport affect the implementation and cancellation

of the flight.

Findings

Based on historical flight information of military aircraft and weather information of flight zone. We developed flight prediction model using data mining techniques. As a result of optimal flight prediction model development for each airbase, it was confirmed that the LightGBM algorithm had the best prediction rate in terms of recall rate. Each flight rules were checked according to the weather condition, and it was confirmed that precipitation, humidity, and the total cloud had a significant effect on flight cancellation. Whereas, the effect of visibility was found to be relatively insignificant. When a flight schedule was established, the rules will provide some insight to decide flight training more systematically and effectively.

Keyword: Data Mining, Big Data Analysis, Military Aircraft, Flight Scheduling, Meteorological Condition Data, Flight Decision, Aircraft Operation Readiness

* 이 논문은 2021년 8월 17일 접수, 2021년 10월 18일 1차 심사, 2022년 1월 27일 2차 심사, 2022년 9월 23일 게재 확정되었습니다.