

Original Article

<https://doi.org/10.12985/ksaa.2021.29.4.001>
ISSN 1225-9705(print) ISSN 2466-1791(online)

머신러닝(Machine Learning) 기법을 활용한 제주국제공항의 운항 지연과의 상관관계 분석 및 지연 여부 예측모형 개발 - 기상을 중심으로 -

이충섭*, Zin Min Paing*, 여혜민**, 김동신***, 백호중****

Development of a Prediction Model and Correlation Analysis of Weather-induced Flight Delay at Jeju International Airport Using Machine Learning Techniques

Choongsub Lee*, Zin Min Paing*, Hyemin Yeo**, Dongsin Kim***, Hojong Baik****

ABSTRACT

Due to the recent rapid increase in passenger and cargo air transport demand, the capacity of Jeju International Airport has been approaching its limit. Even though in COVID-19 crisis which has started from Nov 2019, Jeju International Airport still suffers from strong demand in terms of air passenger and cargo transportation. However, it is an undeniable fact that the delay has also increased in Jeju International Airport. In this study, we analyze the correlation between weather and delayed departure operation based on both datum collected from the historical airline operation information and aviation weather statistics of Jeju International Airport. Adopting machine learning techniques, we then analyze weather condition Jeju International Airport and construct a delay prediction model. The model presented in this study is expected to play a useful role to predict aircraft departure delay and contribute to enhance aircraft operation efficiency and punctuality in the Jeju International Airport.

Key Words : Artificial Intelligence(인공지능), Machine Learning(머신러닝), Decision Tree(결정 나무), Random Forest Classification(랜덤 포레스트 분류), Departure Delay(출발지연), Weather-induced Delay(기상요인에 따른 지연), Jeju International Airport(제주국제공항)

1. 서 론

Received: 05. Nov. 2021, Revised: 17. Dec. 2021,

Accepted: 21. Dec. 2021

* 한국항공대학교 항공교통물류학과 박사과정

** 한국항공대학교 항공교통물류학과 학사과정

*** 한국항공대학교 항공교통물류학과 석사과정

**** 한국항공대학교 항공교통물류학과 교수

연락처 E-mail : hojongbaik@gmail.com

연락처 주소 : 경기도 고양시 덕양구 항공대학로 76

1.1 연구 배경 및 목적

1.1.1 연구 배경

1.1.1.1 제주국제공항

제주국제공항은 대한민국 제주특별자치도 제주시 용담2동, 도두동 일대에 위치한 국제공항으로 국내 주요

공항 중에서 인천국제공항 다음으로 이용객이 많은 공항이다. 1958년 1월 제주비행장으로 정식 설립된 이래, 1968년 제주국제공항으로 승격되었고, Table 1 및 Fig. 1과 같이 2013년에 연간 이용객 2천만 명, 2019년에 3천만 명을 넘어서는 등 최근 빠른 속도로 수요 증가를 경험하고 있다. 특히 2009년 이미 국내선 승객의 경우 김포국제공항을 제치고 국내 공항 중 가장 많은 국내선 승객 이용률을 기록하였으며, 화물 수송량도 꾸준히 증가하고 있는 추세이다. 비록 코로나 19사태로 인하여 일시적으로 승객 수송량의 감소를 보이고 있으나 중, 장기적인 관점에서 볼 때 승객 및 화물 운송의 지속적인 증가가 예상되고 있다.

제주국제공항의 경우 두 개의 활주로가 있으나 남북 방향 활주로(6,234 FT)는 길이가 짧아 소형 항공기의 이

용만 가능하고, 주 활주로인 동서 방향의 활주로(10,433 FT)는 제주도의 풍향이 주로 남동, 북서 방향임을 고려할 때 계절성이 강한 측풍(cross wind)과 저고도 돌풍(low level wind shear)으로 항공기 운항에 어려움이 상존하고 있다. 이와 더불어 협소한 주기장 지역(ramp area)으로 인하여 국제민간항공기구(International Civil Aviation Organization, ICAO)²⁾ 기준 날개폭 65m 이상 F급 대형 항공기의 운항에 제약이 있다.

이러한 공항 주변을 둘러싸고 있는 자연적 환경과 공항 활주로, 유도로 등 공항 시설의 물리적 배치에 의한 미래 확장성의 제약은 향후 항공기 운항의 비약적 증가에 따라 예측되고 있는 제주국제공항 수용량의 포화는 물론 계절적으로 수시로 발생하는 악기상에 따른 지연 초래와 같은 부정적인 영향으로 항공기 운항 지연이 더욱 가중될 것으로 예상되고 있다.



Fig. 1. Jeju International Airport demand status¹⁾

1.1.1.2 제주국제공항 지연 현황

최근 국적 저비용항공사의 집중적인 취항으로 인하여 수용량의 요구가 비약적으로 증가하고 있는 제주국제공항은 지연도 정비례로 증가하고 있으며, 지연율도 전체 운항 편수의 13~22%(2015-2019년 한국공항공사 자료 기준)에 달할 정도로 정시 운항에 부정적 영향

Table 1. Passenger and freight statistics at Jeju International Airport

년도	2015	2016	2017	2018	2019
운항(편수)	158,691	172,743	167,280	168,331	175,366
여객(명)	26,237,362	29,707,364	29,604,363	29,455,305	31,316,394
화물(TON)	278,718	271,494	275,129	266,370	258,847

Annual passenger and freight details

년도	운항(편수)			여객(명)			화물(TON)		
	도착	출발	계	도착	출발	계	도착	출발	계
2015	79,336	79,355	158,691	13,121,331	13,116,231	26,237,562	126,669	152,049	278,718
2016	86,316	86,427	172,743	14,846,593	14,860,771	29,707,364	133,100	138,394	271,494
2017	83,642	83,638	167,280	14,793,321	14,811,042	29,604,363	127,567	147,562	275,129
2018	84,156	84,175	168,331	14,726,651	14,728,654	29,455,305	124,372	141,999	266,370
2019	87,669	87,697	175,366	15,657,540	15,658,854	31,316,394	121,197	137,649	258,847
합계	421,119	421,292	842,411	73,145,436	73,175,552	146,320,988	632,905	717,653	1,350,557

출처: 한국공항공사 자료, 2020.

1) 매일경제 보도자료, 2015년 11월 10일.

2) 국제민간항공기구(ICAO)는 캐나다 소재의 민간항공기구로서, 200여 개의 나라가 모여 향후 국제적인 항공정책 및 규제를 결정하는 조직이다.

을 미치고 있고, 이러한 지연의 원인은 Table 2와 같이 공항기상, 항공기 연결 및 정비, 여객 처리, 복합원인, 기타 등으로 분석되고 있다. 이 중 운항 기상환경과 관련된 지연은 그 비중과 횟수가 점차 증가하는 추세이다.

전 세계적인 기후변화로 인한 기상 현상의 변화는 제주국제공항의 항공기 지연에도 영향을 미치고 있고, 그 빈도와 심도가 점점 예측하기 힘든 형태로 나타나고 있어 승객 및 화물 화주들의 피해가 점차 증가하고 있는 추세이다. 이에 따라 취항 항공사 및 공항 당국의 지연에 대비한 신속한 대응책이 요구되고 있으며, 특히 급변하는 기상 현상에 따른 지연 예측 프로그램 개발의 중요성이 절실히 요구되고 있다.

1.1.2 연구 목적 및 방법

본 연구의 목적은 제주국제공항의 최근 5년(2015-2019년)간의 운항 및 기상 DATA의 통계학적 분석을 통하여 기상요인과 운항 지연의 연관성을 분석하고, 머

신러닝(machine learning) 기법을 활용하여 기상요인을 고려한 운항 지연 예측모형을 개발하는 데 있다.

본 연구에서는 제주국제공항의 최근 5년간(2015-2019)의 운항 및 기상 DATA 분석을 통하여 제주 기상의 계절적, 특징적 FACTOR를 분석, 고찰하고 그 결과를 활용하여 머신러닝 기반의 지연 예측모형으로 구축하였다. 구축된 예측모형은 악기상이 예상되는 날의 지연 여부를 예측하고, 이를 토대로 운항 기재변경, 운항 스케줄 조정 등의 정책적 대응 방법을 통해 항공사의 지연 최소화 및 공항 운영의 효율성 확보에 도움을 주고자 한다.

1.1.3 연구 범위

본 연구의 공간적 범위는 제주국제공항이며, 시간적 범위는 코로나 19로 인한 항공 교통량의 왜곡이 생긴 2020년을 제외한 최근 5년(2015-2019년)간으로 설정하였다. 연구의 내용적 범위는 제주공항 출발 항공기 총 422,218편에 대해 최근 제안된 항공기 운항 지연

Table 2. Jeju International Airport delay statistics

년도	출/도착	운항 (편수)	지연(편수)						계
			기상	A/C 접속	A/C 정비	여객 처리	복합 원인	기타	
2015	출발	77,254	117	14,750	93	58	0	122	15,140
	도착	77,321	258	4,478	112	39	2	204	5,093
	계	154,575	375	19,228	205	97	2	326	20,233
2016	출발	84,467	136	26,396	87	38	0	142	26,799
	도착	84,910	321	9,730	106	90	3	353	10,603
	계	169,377	457	36,126	193	128	3	495	37,402
2017	출발	81,862	179	16,572	131	69	0	170	17,121
	도착	81,816	265	4,856	122	38	0	250	5,531
	계	163,678	444	21,428	253	107	0	420	22,652
2018	출발	82,098	254	19,060	202	107	2	238	19,863
	도착	82,190	356	5,934	105	42	0	195	6,632
	계	164,288	610	24,994	307	149	2	433	26,495
2019	출발	85,389	280	17,354	180	141	2	191	18,148
	도착	85,465	430	5,021	162	65	5	221	5,904
	계	170,854	710	22,375	342	206	7	412	24,052

출처: 한국공항공사 자료, 2020.

기준안을 적용하여, 지연의 주요 원인 중 하나인 기상 관련 지연 여부 판별 및 예측 모델(prediction model) 구성으로 제한하여 진행하였다.

항공 기상청을 통해 수집된 기상 분석 전문(meteorological terminal air report, METAR) 내용 중 주요 변수 중 7가지를 선정하여 지연과의 상관관계를 분석하였고, TS(thunderstorm), WS(wind shear)의 유무에 따른 지연 여부를 분석하였다. 분석에 사용한 TOOL은 PYTHON 프로그램이 사용되었으며, 주어진 DATA들의 분석 기법은 머신러닝 기법 중 지도 학습(supervised learning) 분야의 랜덤 포레스트 분류(random forest classification) 기법을 적용하였다.

II. 본 론

2.1 기존 문헌 및 분석 기법 고찰

2.1.1 항공기 지연

2.1.1.1 지연의 정의

본 연구의 핵심적인 대상인 항공기 운항 지연의 국제적 기준은 〈운항 시각표+15분〉을 기준으로 지연 여부를 판단하여 통계치를 산정하고 있다. 이와는 달리, 국내에서는 비행 계획서상의 운항 시간에서 이, 착륙 시간을 기준으로 국내선은 30분 초과, 국제선은 1시간 초과한 경우 지연으로 간주하여 통계를 작성³⁾하고 있다. 한편, 항공사의 정식 운항 기준은 항공사에 따라 스케줄 출발시각 대비 0분, 5분 및 15분 이내 출발 시 정식 운항으로 간주하는 등 자체 기준을 설정하여 운영하고 있다. Ryu(2008)는 미국 연방항공청(Federal Aviation Administration, FAA)⁴⁾ 기준을 반영하여 출발 정시율은 〈STD+15분〉 이내, 도착 정시율은 〈STA+15분〉 이내를 기준으로 연구를 수행하였다. 미국의 경우 주요 항공사들로부터 지연 원인을 포함한 운항 자료를 접수하여 매월 미국 교통부(Department of Transportation, DOT) 산하기관인 교통 통계국(Bureau of Transportation Statistics, BTS)을 통해서 통계를 공시하고 있다.

그러나, 각 지연의 기준을 살펴보면 실제 출발시각(actual time of departure, ATD)과 실제 도착시각(actual time of arrival, ATA)의 기준이 상이하다(Kim et al., 2019). 예컨대, 우리나라 항공통계는 출발 항공기가 활주로에서 이륙 후, 감시레이더에 최초로 포착되는 시각을 기준으로, 항공사는 항공기가 주기장에서 이동을 시작하는 시각(actual off-block time, AOBT)을 기준으로, 미국 연방항공청은 항공기가 브레이크를 해제(release)하는 시각(taxi-out time)을 기준으로 실제 출발시각을 기록한다. 또한, 도착 항공기에 대해 우리나라 항공통계는 항공기가 활주로에 접지하는 시각을, 항공사는 항공기가 주기장에 도착하는 시각(actual in-block time, AIBT)을 기준으로, 미국 연방항공청은 지상주행 시각(taxi-in time)을 기준으로 실제 도착시각을 적용하고 있다.

2.1.2 머신러닝(Machine Learning)

2.1.2.1 머신러닝 개요

최근 컴퓨터 성능의 비약적인 발전과 함께 지속적으로 진화하고 있는 분야인 인공지능(artificial intelligence, AI)⁵⁾은 인간의 지능으로 할 수 있는 사고, 학습, 자기계발 등을 컴퓨터가 학습할 수 있게 하는 방법을 연구하는 컴퓨터공학 및 정보 기술의 한 분야이다. 즉, 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것을 말한다.

머신러닝⁶⁾(또는 기계학습)이란 컴퓨터 과학 중 인공지능의 한 분야로 패턴인식과 컴퓨터 학습 이론의 연구로부터 진화한 분야이다. 머신러닝은 경험적 데이터를 기반으로 학습을 하고 예측을 수행하며 스스로 성능을 향상하는 시스템과 이를 위한 알고리즘을 연구하고 구축하는 기술이라 할 수 있다. 엄격하게 정해진 정적인 프로그램 명령을 수행하는 것보다는 입력 데이터를 기반으로 예측이나 결정을 도출하기 위해 특정한 모델을 구축하는 방식을 취한다. 실제로 인간이 프로그래밍을 통해 필요한 모든 단계를 설정하는 것보다 기계가 자체 알고리즘을 개발하도록 돕는 것이 더 효과적이다. 머신러닝은 기존에 사용된 알고리즘들로 개발

3) 출처: 국토교통부, 항공통계 작성 매뉴얼, 2013.

4) 미국 연방항공청은 미국 내의 민간항공 안전을 위한 주요 임무를 위해 설립된 정부 기관이다.

5) 본 논문에서 소개하는 인공지능 관련 내용은 아래 주소의 내용을 분류, 정리한 것임, <https://terms.naver.com/entry.nhn?docId=3386834&cid=58369&categoryId=58369>

6) 본 논문에서 소개하는 머신러닝 방법론 내용은 아래 주소의 내용을 분류, 정리한 것임, https://en.wikipedia.org/wiki/Machine_learning

하기 어렵거나 실행 불가능한 이메일 필터링 및 컴퓨터 비전과 같은 다양한 애플리케이션에서 사용되고 있다. 머신러닝은 학습 시스템에서 사용할 수 있는 “신호” 또는 “피드백”의 특성에 따라 3가지 유형으로 구분된다.

- 1) 지도학습: 입력과 대응하는 미리 알려진 출력(인간 전문가가 제공)을 매핑(mapping)하는 함수를 학습하는 과정으로, 크게 분류와 회귀 문제를 해결하는 형태로 나뉜다.
 - 분류(classification): 결과 출력이 “예” 또는 “아니오”, “참” 또는 “거짓”, “0” 또는 “1”과 같은 범주형(categorical)인 문제를 의미함.
 - 회귀(regression): 연속적(continuous)인 출력값을 예측하는 문제를 의미함.
- 2) 비지도 학습(unsupervised learning): 출력 없이 입력만으로 모형을 구축하여 학습함. 일반적으로 데이터 마이닝의 대부분 기법이 해당함.
 - 군집화(clustering): 데이터의 특성을 분석해 관련 있는 데이터들을 묶어주는 문제를 의미함.
- 3) 강화학습(reinforcement learning): 학습자가 행동을 수행함으로써 환경에 영향을 미치고, 이에 대한 피드백으로 보상을 얻어 학습 알고리즘의 가이드로 사용하는 문제를 의미함.

2.1.2.2 결정 나무(Decision Tree)

Fig. 2는 결정 나무(decision tree)의 개념을 도식화한 것으로 노드와 분기로 구성된 순서도(flow chart)와 같은 구조로 이루어져 있다. 각 노드에서 입력 변수

Decision Tree

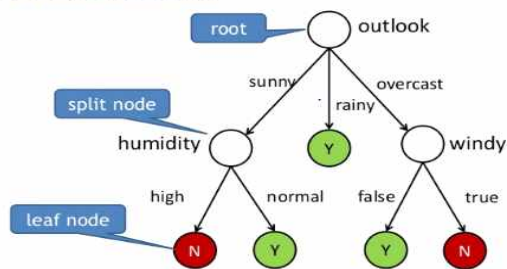


Fig. 2. Decision tree structures⁷⁾

중 하나를 기반으로 데이터 분할이 수행되고, 두 개의 이상의 분기를 출력으로 생성한다. 다가오는 노드에서 점점 더 많은 분할이 이루어지고 원본 데이터를 분할하기 위해 점점 더 많은 분기가 생성된다. 이것은 모든 또는 거의 모든 데이터가 동일한 클래스에 속하고 더 이상의 분할 또는 분기가 불가능한 노드가 생성될 때까지 계속된다.

전체 과정은 나무와 같은 구조를 생성한다. 첫 번째 분할 노드를 루트 노드(root node), 중간 노드를 내부 노드(split node), 끝 노드를 종단 노드(leaf node)라고 한다. 종단 노드는 분류(CLASS)의 LABEL을 의미하며, 규칙에 따라 루트에서 종단 노드까지의 경로가 생성된다.

2.1.2.3 랜덤 포레스트 분류(Random Forest Classification)

랜덤 포레스트 분류 알고리즘은 지도학습의 분류 문제를 해결하기 위한 머신러닝 모형 중 하나로서 의사결정 나무 모형의 진화된 버전이라고 할 수 있다. 머신러닝에서 랜덤 포레스트 모형이란 분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로, Fig. 3과 같이 훈련 과정에서 구성한 다수의 결정 나무로부터 부류(분류) 또는 평균 예측치(회귀 분석)를 출력함으로써 동작한다. 랜덤 포레스트 모형의 장점은 많은 결정 나무가 하나의 단일 결정 나무보다 더 정확한 예측을 할 수 있다는 것이다. 실제로 랜덤 포레스트 분류 알고리즘은 N개의 서로 다르게 훈련된 결정 나무를 구축하고 이를 병합하여 더 정확하고 안정적인 예측을 얻는 감독된 분류 알고리즘이다.

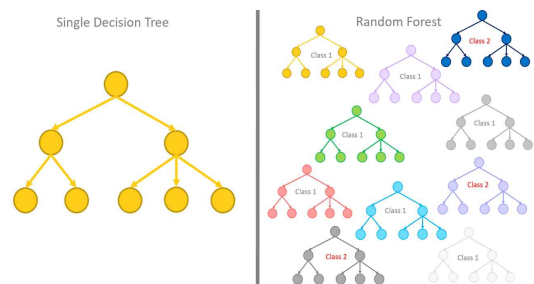


Fig. 3. Single decision tree and random forest structures⁸⁾

7) Genesis, “Decision Tree”, <https://www.fromthegenesis.com/decision-tree>

8) Silipo, R. and Melcher, K., “From a Single Decision Tree to a Random Forest”, 2019.

항공기 지연의 정의에 대한 기존 문헌 고찰을 통해, 본 연구에서는 항공기 지연 중 출발지연을 Table 3과 같이 정의하였다. 항공기가 이륙하는 시각을 출발에 대한 기준시각으로 제시하고, 출발 항공기의 지상 주행(taxi-out) 시간(15분)을 고려하여 항공기 지연 여부의 판별기준(30분)을 설정하였다. 또한, 머신러닝 기법 중 랜덤 포레스트 분류 알고리즘을 사용하였는데, 이는 랜덤 포레스트 모형이 다른 모형과 비교할 때 다량의 데이터를 안정되게 처리할 수 있기 때문이다.

2.2 연구 방법

2.2.1 자료 수집 및 전처리

2.2.1.1 자료 수집: 통계자료

본 연구에 활용된 자료는 크게 1) 제주공항 2015-2019년 항공기 운항 정보와 2) 2015-2019년 항공 기상 통계자료이다. 항공기 운항 관련 정보는 국토교통부 항공정보시스템(UBIKAIS)에서 제공하는 항공기 운항 및 비행정보시스템(Flight Operation Information System, FOIS) 자료를 사용하였으며, 항공 기상 통계 자료는 항공 기상청에서 제공하는 METAR 및 SPECI (SPECIal Reports) 자료를 사용하였다.

Table 3. Aircraft departure delay criteria

지연기준(분)	출발시각 기준
$ATD - STD \geq 30$	할주로 이륙시각

FOIS 자료는 항공기 호출부호(callsign), 항공기 기종(aircraft type), 항공기 등록 번호(aircraft registration number), 운항 스케줄 상의 출발, 도착공항(origin/destination airports), 운항 날짜(date), 운항 스케줄 상의 출발, 도착시각(scheduled time of departure/arrival), 실제 출발, 도착시각(actual time of departure/arrival) 등의 정보를 포함하고 있다. METAR 자료는 협정 세계시각(UTC time), 풍향(wind direction), 풍속(wind speed), 시정(visibility), 일기 현상(significant weather), 운고(cloud height), 운량(cloud amount), 기온(temperature), 기압(atmosphere) 등의 정보를 포함한다. Fig. 4와 Fig. 5는 각각 항공기 운항 정보와 항공 기상 통계자료의 예시를 보여준다.

2.2.1.2 수집자료 전처리

신뢰성 있는 연구결과를 도출하기 위해서는 입력자료의 정확성이 요구된다. FOIS 자료의 경우 일부 정보가 누락 및 신뢰도가 낮은 항공편의 기록은 분석에서 제외하였다. 데이터 처리 과정을 거친 결과 얻은 2015-2019년간 제주공항 출발 항공편은 총 426,704편에서 4,486편이 제거된 총 422,218편이다. 또한, METAR 자료에서 일부 누락된 시간대의 데이터에 대해 해당 시간대의 직전 또는 직후 데이터를 사용하였다.

Fig. 4 및 5는 각각 전처리 전의 FOIS와 METAR 자료를 예시한 것으로 전처리 과정을 거친 후 FOIS와 METAR 자료를 날짜(date)와 시각(hour)에 대해 병합하였다. Table 4는 변수(feature)와 각 변수의 형태를

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
flightPK	callsign	actype	registration	orgApt	schDate	std	etd	atd	desApt	staDate	sta	eta	sta	status	fpl/yln	blockTime	stanc
25490271	KEA361	E145	HL8054	RKNV	20190401	0950	0950	0950	RKPC	20190401	1125	0	1100	ARR	0	0	064
25490891	KEA363	E145	HL8054	RKNV	20190401	1700	0	1655	RKPC	20190401	1820	0	1802	ARR	0	0	064
25635776	PIC1034	A320	VNA565	VVNB	20190401	0015	0215	0224	RKPC	20190401	0605	0543	0537	ARR	Y	0	015
25703172	ESR552	B738	HL8058	VIBS	20190401	0225	0425	0507	RKPC	20190401	0920	0934	0933	ARR	Y	0	018
25703312	ESR884	B738	HL8343	RCTP	20190401	1115	1220	1244	RKPC	20190401	1435	1429	1419	ARR	Y	0	017
25709942	CQ48567	A320	B6902	ZSPD	20190401	1805	1905	1906	RKPC	20190401	2035	2003	2011	ARR	Y	0	018
25711465	CQ48913	A320	B1671	ZBTI	20190401	1935	2035	2037	RKPC	20190401	2230	2216	2213	ARR	Y	0	018
25713276	CE55059	A321	B8573	ZSPD	20190401	0945	1045	1117	RKPC	20190401	1230	1217	1226	ARR	Y	0	020
25714540	TWB246	B738	HL8324	RUBB	20190401	1255	1255	1304	RKPC	20190401	1505	1445	1436	ARR	Y	0	018
25714990	UA2186	B738	HL8034	VHMH	20190401	0225	0325	0342	RKPC	20190401	0610	0615	0603	ARR	Y	0	053
25715174	KAL718	B738	HL2011	RUAA	20190401	0940	0940	0951	RKPC	20190401	1225	1239	1219	ARR	Y	0	015
25715509	KAL880	A333	HL7587	ZBAA	20190401	1055	1155	1214	RKPC	20190401	1425	1408	1402	ARR	Y	0	032
25717405	CSN6097	A320	B9911	ZYTL	20190401	1855	1955	1950	RKPC	20190401	2145	2115	2106	ARR	Y	0	031
25718002	JNA102	B738	HL7555	ZSPD	20190401	0710	0810	0823	RKPC	20190401	0940	0926	0922	ARR	Y	0	017
25718272	JNA172	B738	HL8015	ZLXY	20190401	0225	0325	0339	RKPC	20190401	0620	0629	0622	ARR	Y	0	017
25718459	KAX501	A333	9M-KXU	WMAK	20190401	0650	0750	0801	RKPC	20190401	1400	1335	1331	ARR	Y	0	020
25718700	OJA2717	B739	B1420	ZBTI	20190401	0845	1000	1000	RKPC	20190401	1200	1140	1143	ARR	Y	0	017
25947090	FEA722	MD83	B28025	RCTP	20190401	1840	1930	1955	RKPC	20190401	2135	2133	2135	ARR	Y	0	057
25947679	DH1375	A320	B304G	ZSPD	20190401	0855	0855	0904	RKPC	20190401	1035	1017	1007	ARR	Y	0	020
25954739	EPAA6329	B738	B1159	ZSNT	20190401	1720	1820	1821	RKPC	20190401	2005	1942	1936	ARR	Y	0	050
25956971	JNA301	B772	HL7743	RKSS	20190401	0605	0605	0622	RKPC	20190401	0720	0710	0710	ARR	Y	0	013
25957181	JNA302	B738	HL7561	RKSS	20190401	0610	0610	0627	RKPC	20190401	0725	0715	0715	ARR	Y	0	056
25957390	JNA305	B738	HL8242	RKSS	20190401	0705	0705	0721	RKPC	20190401	0820	0810	0811	ARR	Y	0	057
25957657	JNA309	B738	HL7561	RKSS	20190401	0950	0950	1010	RKPC	20190401	1105	1055	1107	ARR	Y	0	057
25957867	JNA311	B772	HL7743	RKSS	20190401	1025	1025	1035	RKPC	20190401	1135	1130	1133	ARR	Y	0	031
25958079	JNA313	B738	HL8243	RKSS	20190401	1055	1055	1108	RKPC	20190401	1205	1200	1211	ARR	Y	0	009
25958233	JNA317	B738	HL8242	RKSS	20190401	1520	1520	1544	RKPC	20190401	1630	1625	1631	ARR	Y	0	001
25958443	JNA319	B738	HL7561	RKSS	20190401	1335	1335	1359	RKPC	20190401	1450	1440	1447	ARR	Y	0	002
25958653	JNA321	B772	HL7743	RKSS	20190401	1440	1440	1459	RKPC	20190401	1555	1545	1553	ARR	Y	0	031

Fig. 4. FOIS data structures

Fig. 5. METAR data structures

Table 4. Variable details

변수 종류	변수(Features)	Remark	예시
비행정보 데이터(입력변수)	Day of the month	범주	1(1일), 2(2일), ..., 31(31일)
	Day of the week	범주	1(일), 2(월), ..., 7(토)
	Hour of the day	범주	1, 2, ..., 24
기상 데이터(입력변수)	Wind direction	범주	0(0도), 1(10도), ..., 36(360도)
	Wind speed	연속	연속값
	Visibility	연속	연속값
	Significance weather	범주	1(RN), 2(SN), ..., 등
	Cloud height	연속	연속값
	Cloud amount	범주	1(FEW), 2(SCT), ..., 등
범주(출력변수)	1 if Delay \geq 30 Mins 0 if Delay < 30 Mins		-

정리한 것이다. 일(day of the month), 요일(day of the week), 시각(hour of the day), 풍향, 풍속, 시정(visibility), 일기현상, 운고, 운량, 그리고 출력 변수인 지연 여부("0" 또는 "1")를 포함한다.

2.2.2 모형 구축

본 연구에서는 제주공항에서 출발하는 항공편의 지연 여부를 예측하기 위해 FOIS 자료와 METAR 자료를 사용하여 항공기 출발지연 예측모형을 구축하였다. 모형 구축과정은 1) 제주공항 월별 지연 비율 분석을 통한 분석 기간 설정 2) 선정한 분석 기간에 대해 지연 유발 원인 분석 3) PYTHON 프로그래밍 툴을 활용한 랜덤 포레스트 예측모형 구축 4) 예측모형의 정확도 평

가의 순서로 이루어져 있다. Fig. 6은 모형을 구축과정 을 순서도 형태로 표현한 것이다.

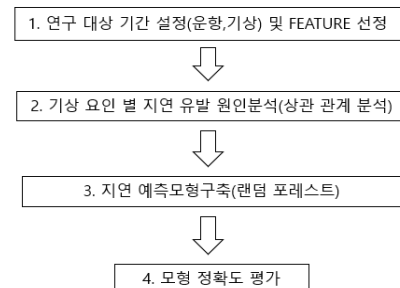


Fig. 6. Process of building predictive models

2.2.2.1 연구 대상 기간 및 FEATURE 설정

연구 대상 기간은 2015년 1월부터 2019년 12월까지로 제주공항의 월별 지연 항공기 편수와 지연율을 분석하였고, 분석을 위한 FEATURE는 마찬가지로 동기간 동안의 기상 정보 중 일(day of the month), 요일(day of the week), 시각(hour of the day), 풍향, 풍속, 시정, 일기현상, 운고, 운량, 그리고 출력 변수인 지연 여부 등이다.

2.2.2.2 출발지연 원인과 기상변수 상관관계 분석

Fig. 7은 기상 DATA를 기준으로 출발지연과 기상변수 간의 상관관계를 분석한 결과이다. 세로축과 가로축의 year, month, day, hour, STD, ATD, WD, WSPD(wind speed), GUST, VIS, WS_RWY(wind shear), TS(thunder storm)는 출발지연에 영향을 주

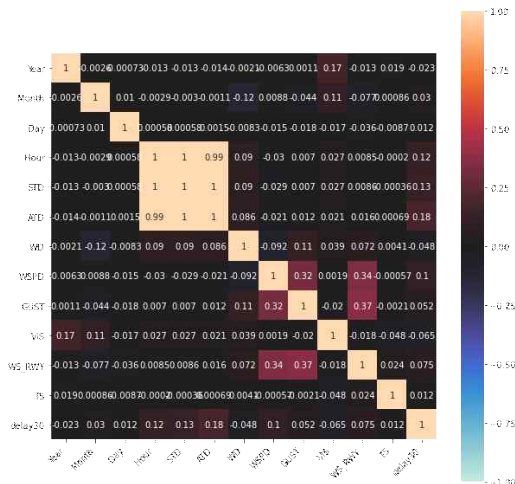


Fig. 7. Correlation between departure delay and variable diagram

는 독립변수(원인)를 나타낸 것이고, delay 30은 출발 항공기의 지연 여부를 나타내는 종속변수(결과)이다. 두 변수가 교차하는 점의 값은 해당 변수들의 상관계수이다. 이때, '+1(-1)' 값은 두 변수 사이에 완벽한 양(음)의 상관관계가 있음을 의미하고, 0 값은 두 변수 사이에 상관이 없음을 의미한다. 분석결과 WSPD, WS_RWY, GUST 변수는 서로 어느 정도 상관관계가 있음을 알 수 있었으며, ATD와 출발지연 간에도 어느 정도 상관관계가 있음을 파악할 수 있다.

2.2.3 지연 분석

연구 대상 총 422,218편 중 지연 운항된 편수는 104,593편으로 분석되었으며, 이는 대상 편수의 약 24.8%에 해당한다. Fig. 8 및 9는 이를 년도 별, 개별 달로 분석하여 5년간 비올과 편수로 분류하여 표로 도식화하여 정리한 것이다. 분석 결과, 계절, 년도, 월 등 기준에 따라 일정한 비올이나 정비례 관계 등 지연을 설명할 수 있는 근거는 찾기 어려웠다. 다만 특징적으로 2016년은 전년 대비 LCC 항공사의 취항 편수가 급증함에 따라 지연율도 동반 상승하였으며, 2017년 고속탈출유도로(high speed taxiway)의 건설과 함께 확보된 공항 효율성에 따라 이후 연도의 지연율은 다소 감소 추세를 보였다.

2.2.4 기상 요소별 분석

2.2.4.1 강수 현상(Precipitation)

비, 눈 등으로 인한 미끄러운 활주로에서는 활주로 마찰 계수가 낮다. 불충분한 활주로 마찰력은 항공기의 활주로 과주(overrun) 및 이탈의 주된 요인이며, 활주로 표면의 마찰력이 감소하면 측풍 허용기준이 감소한다. 또한, 활주로의 제동능력의 감소로 인하여 항공기

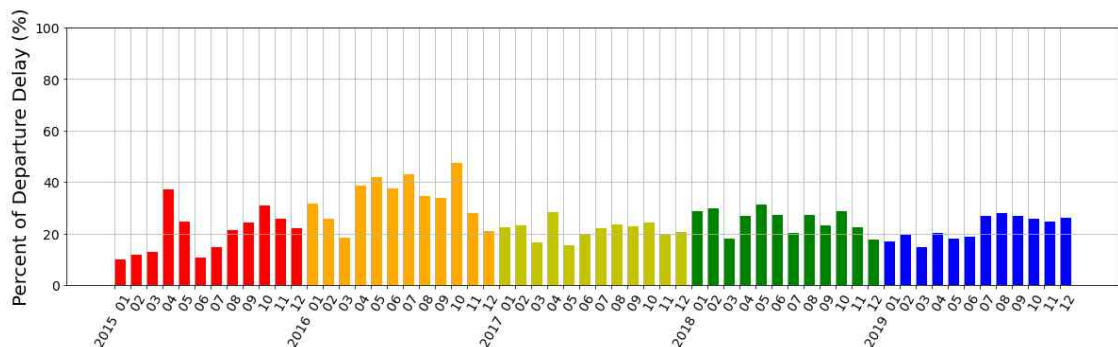


Fig. 8. Number of departure delays by month (2015-2019)

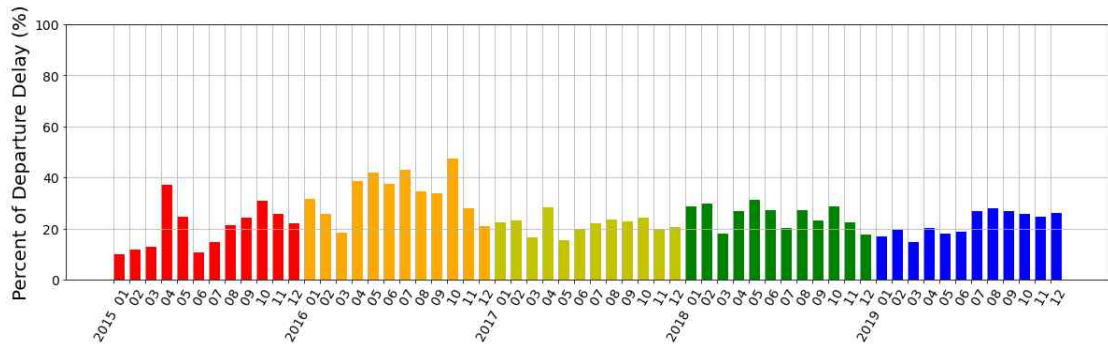


Fig. 9. Percent of departure delay by month (2015-2019)

의 활주로에서의 머물게 되는 시간 즉, ROT(runway occupation time)가 늘어나게 되어 전체적인 활주로 수용량에도 부정적인 영향을 미친다. 활주로 마찰 계수는 항공기 운항의 정규성과 효율성에 매우 중요하다. 강수 현상별로 지연과의 상관관계를 분석하면 총 지연 출발된 104,593편 중 실질적으로 강수 현상과 관련된 출발지연은 총 16,878편으로 Table 5와 같이 적지 않은 영향을 미치는 것으로 분석되었다.

2.2.4.2 시정(Visibility)

시정 기준 1,000M 이하, 즉 FG 수준을 하회하는 상황에서 출발한 항공편수는 5년간 총 출발편의 미미

Table 5. Analysis between departure delay and precipitation and weather variables

(Unit: Number of departure delays)

	2015년	2016년	2017년	2018년	2019년
-RA, RA, +RA, -SHRA, SHRA	2,478	4,671	1,627	2,341	3,294
-SN, SN, +SN, -SHSN, SHSN	144	230	377	1,137	10
-DZ, DZ	163	191	29	77	109
BR, FG, DU, HZ	1,574	1,593	612	461	695
WIND-SHEAR	1,725	3,263	2,757	2,997	2,038
THUNDER- STORM	0	38	48	19	74

주1: RA(rain), SHRA(shower rain), SN(snow), SHSN (shower snow), DZ(drizzle), BR(mist), FG(fog), DU(dust), HZ(haze).

주2: (-): Light, (+): Heavy.

한 수준인 919편이고, 이 중 지연 운항된 편은 513편으로 조사되었다. 제주국제공항의 자연적인 운항환경을 고려하면 실질적으로 저시정 상황에서 운항되는 경우는 상대적으로 미미한 수준이고, 기타 BR, DU, HZ 등의 상황을 고려하더라도 시정과 관련된 기상 요인이 실질적인 지연의 주요한 원인으로 보기는 어려운 것으로 분석되었다. 그럼에도 불구하고 저시정 상황에서 출발한 919편 가운데 513편(55.8%)이 지연되어 시정은 주의를 요구하는 기상 요소로 분석되었다.

2.2.4.3 바람(Wind)

본 연구를 통하여 제주국제공항의 지연에 가장 큰 영향을 미치는 기상요인이 바람과 관련된 요인임이 파악되었고, 이를 세분화하여 풍속(wind speed), 순간 돌풍(gust), 급변풍(windshear) & 뇌전(thunderstorm), 측풍(crosswind) 등의 부문으로 분류하여 아래와 같이 항목별로 분석하여 보았다. 제주국제공항에서의 바람의 방향과 지연과의 상관관계를 일률적으로 분석하기는 매우 어려웠으며, 또한 연구 대상 기간 동안 월별 풍향과 풍속 분석에서도 동절기를 제외하고는 일정한 경향성을 발견하기가 어려워 월별로 분류하여 Fig. 10-14와 같이 그림으로 표시하였다.

2.2.4.3.1 풍속(Wind Speed)

연구결과, 제주국제공항의 주요 활주로인 동, 서 활주로(RWY 07, 25) 기준으로 항공기 운항이 풍속 기준 10 KT 이상 불고 있는 상태에서 출발 운항된 총 134,737편 중 40,038편이 출발지연 운항된 것으로 파악되었고, 특히 15 KT 이상의 바람이 부는 상황에서 출발 운항한 55,488편 중 19,098편이 지연 운항된 것으로 연구되어 전체 운항 지연 편수 104,593편 중 다른 지연

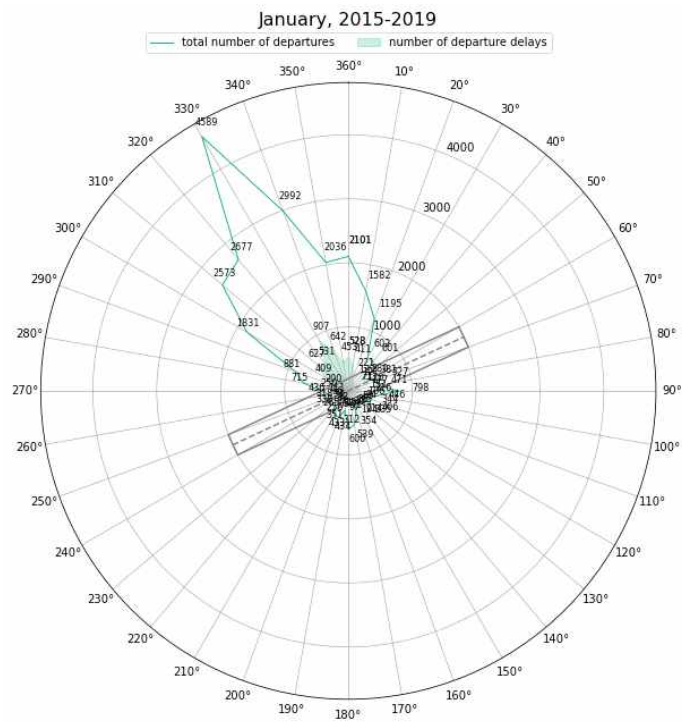


Fig. 10. Number of departure delays with respect to wind directions (Jan., 2015-2019)

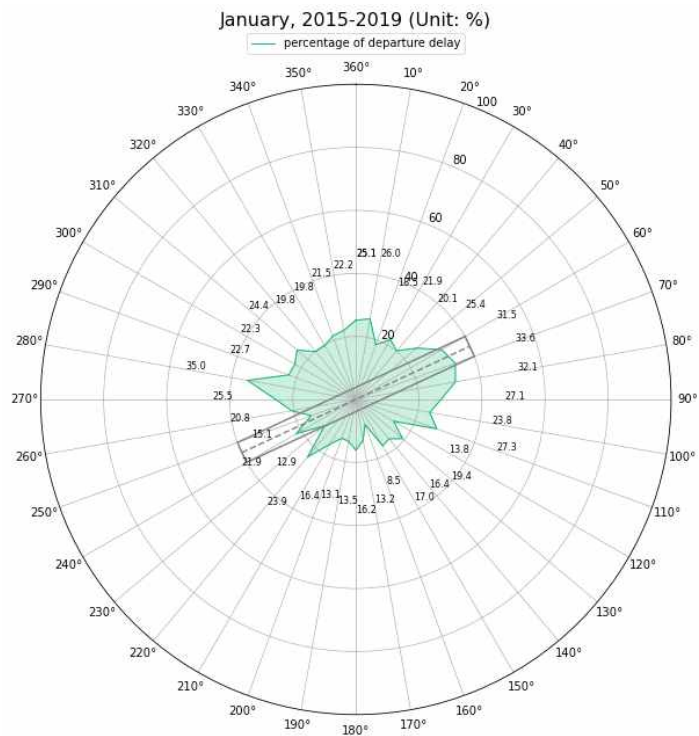


Fig. 11. Percentage of departure delay with respect to wind directions (Jan., 2015-2019)

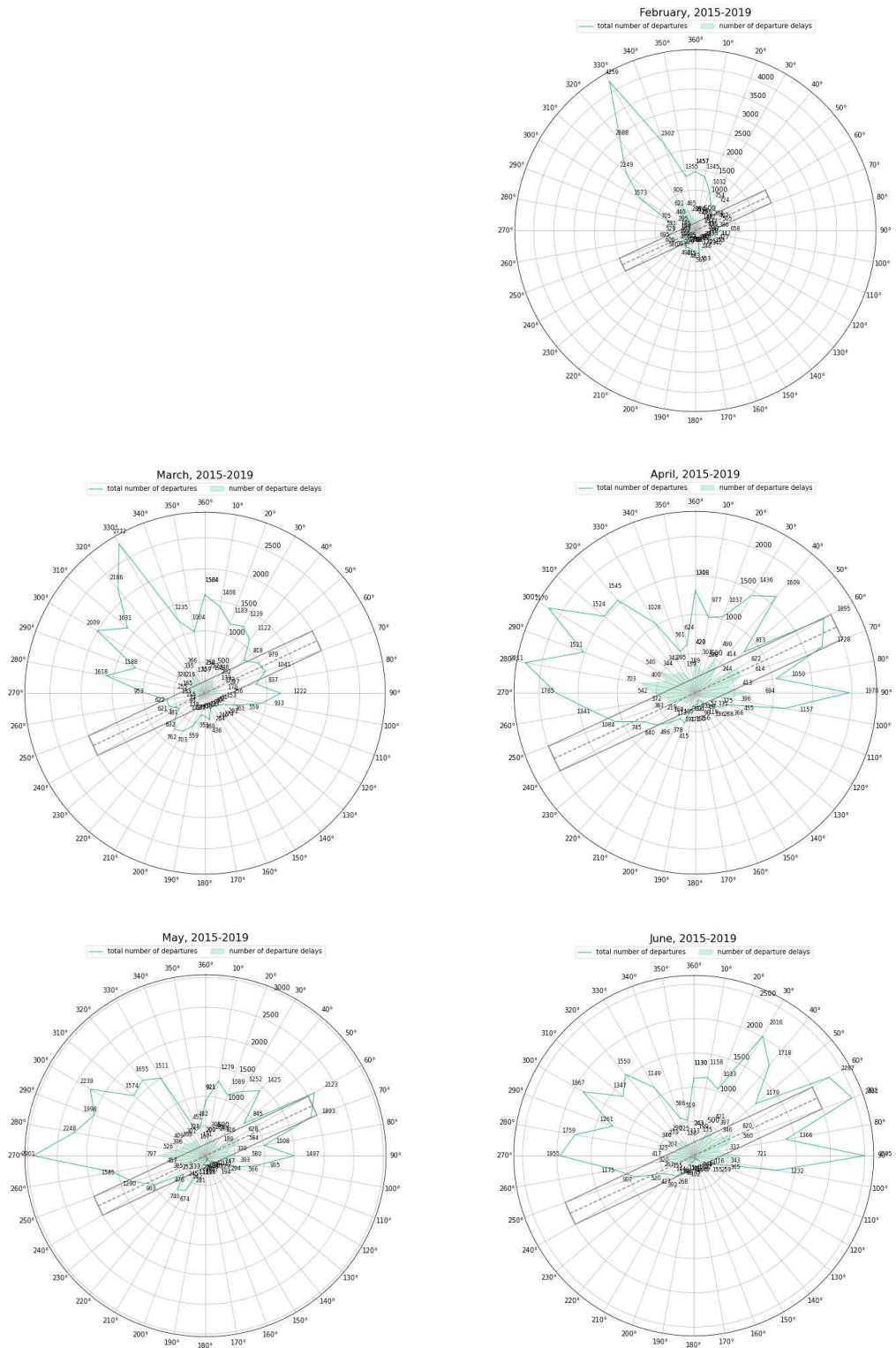


Fig. 12. Monthly number of departure delays by wind directions (Total in 2015-2019)

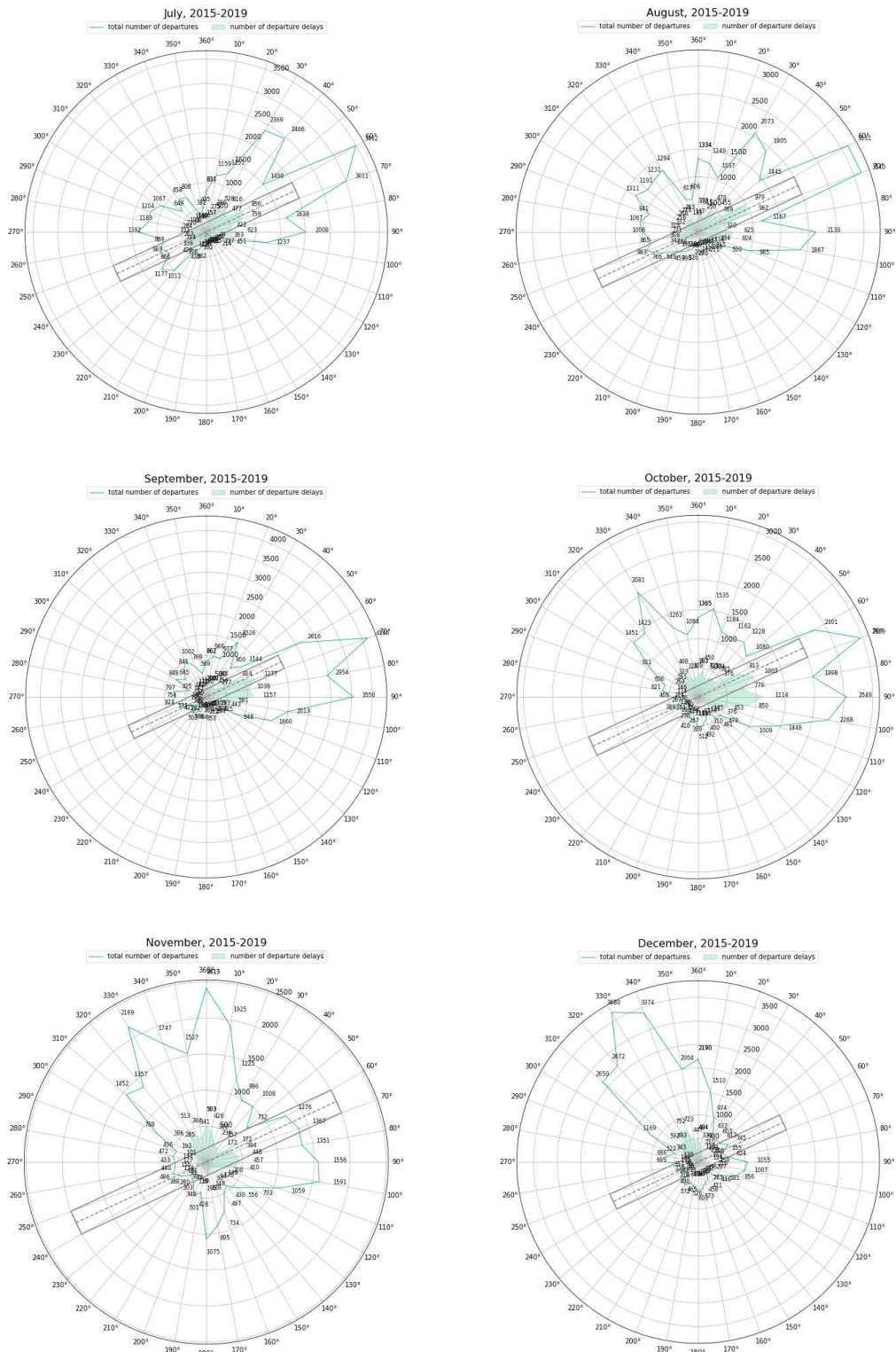


Fig. 12. Continued



Fig. 13. Monthly percentage of departure delay by wind directions (Total in 2015-2019)

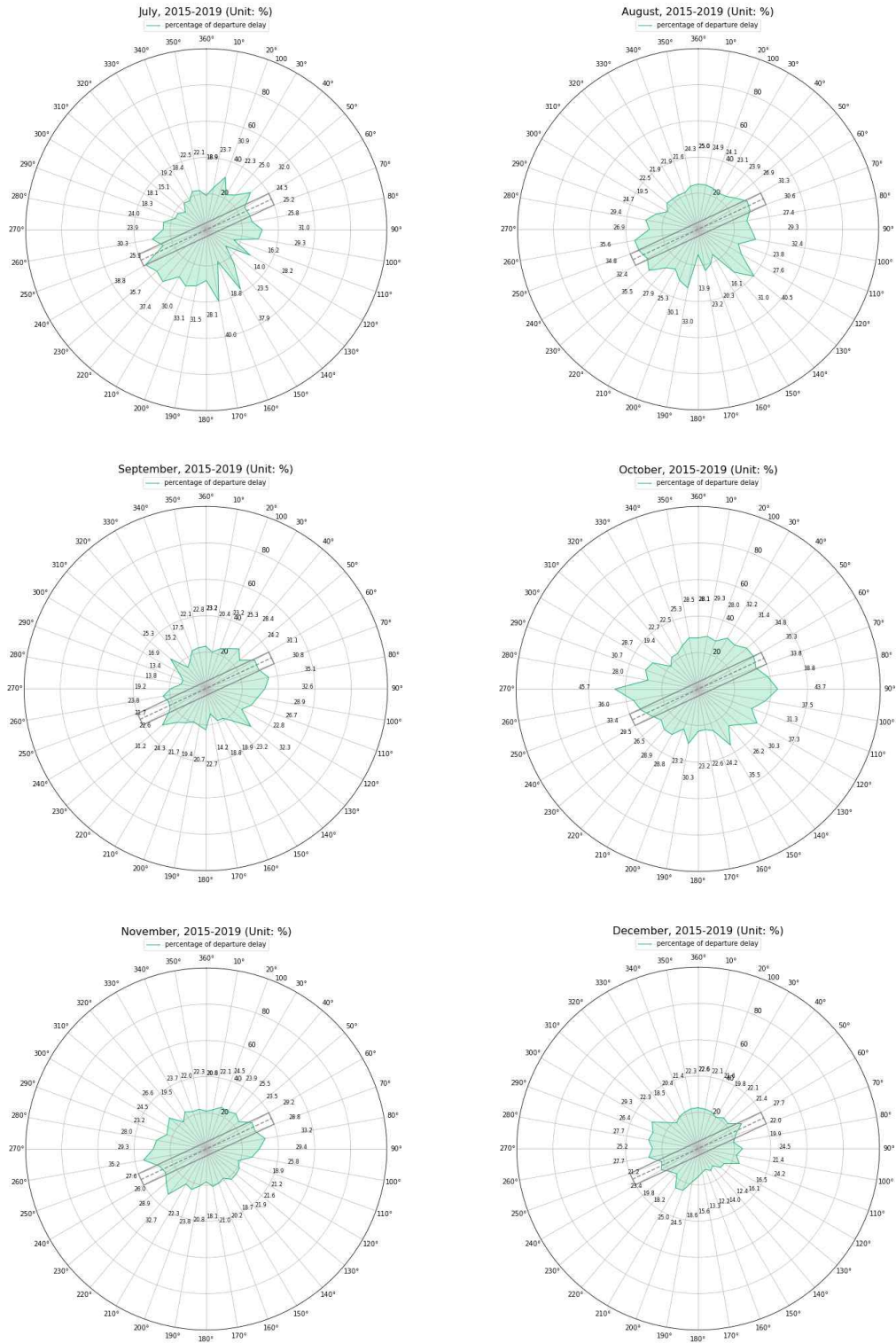


Fig. 13. Continued

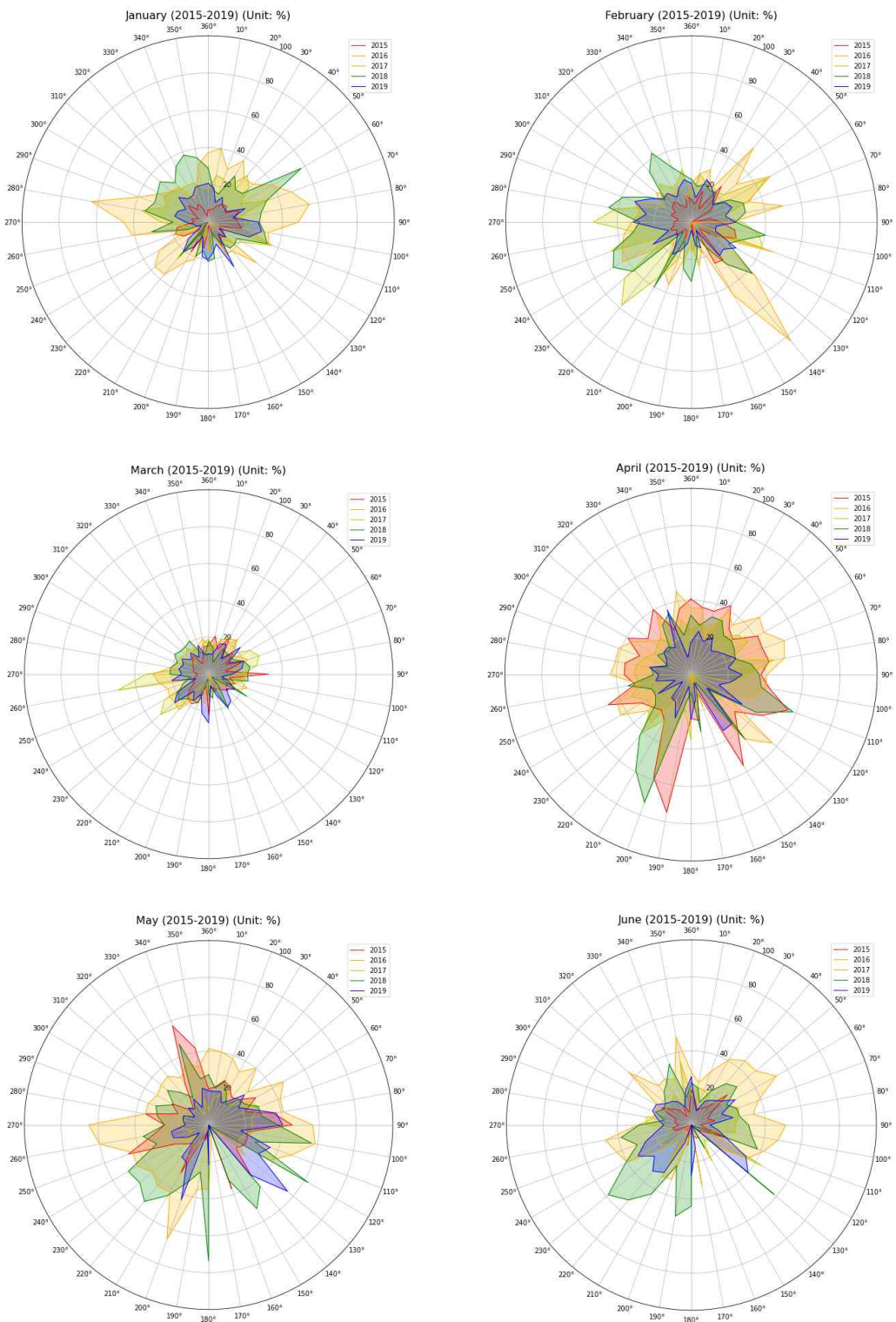


Fig. 14. Monthly percentage of departure delay by wind directions (2015-2019)

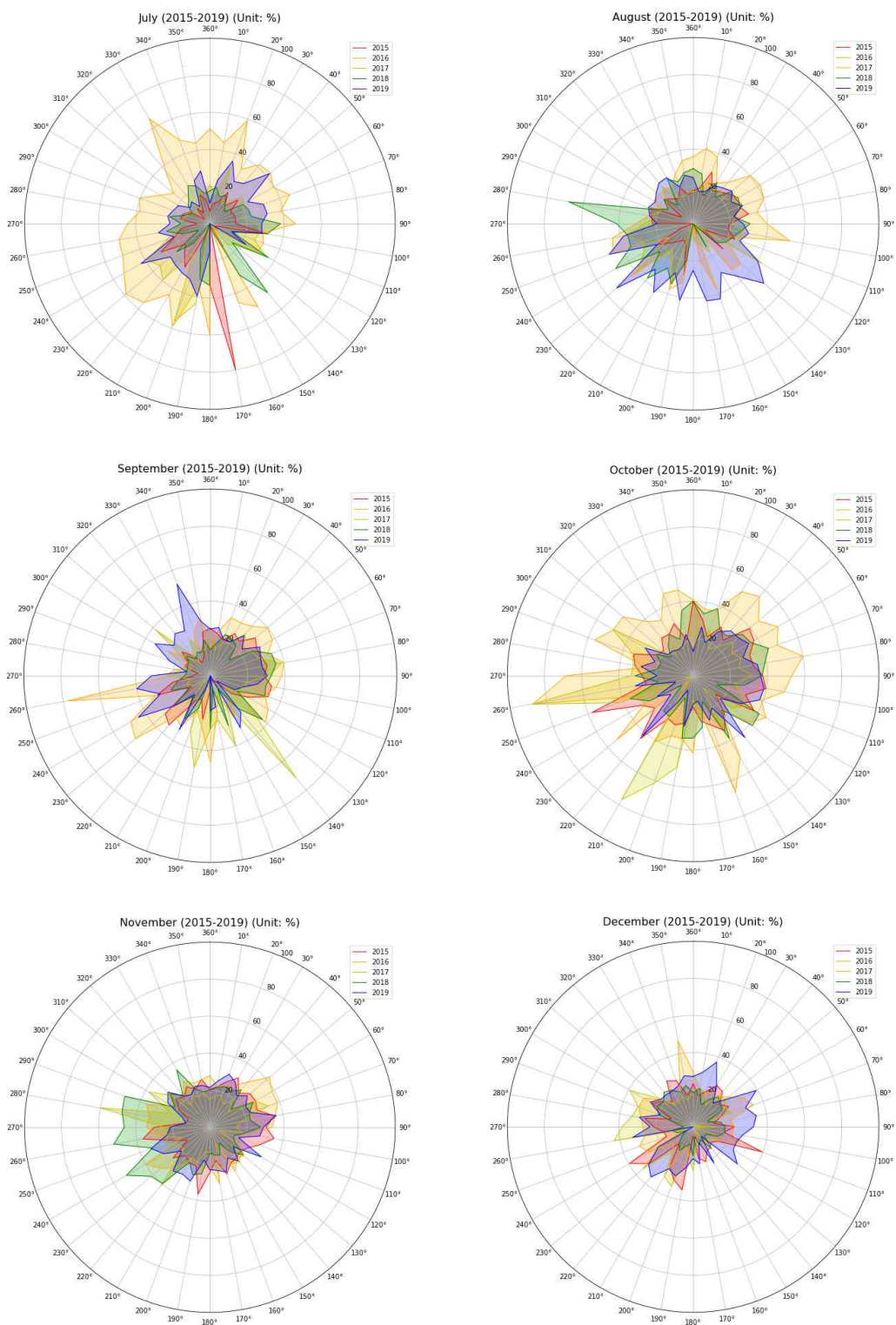


Fig. 14. Continued

적으로 예약된 원본 데이터의 20%를 랜덤 포레스트 분류기를 통해 구축된 모형에 입력 데이터로 사용하였다. 모형의 정확도 테스트 결과, 5년간의 데이터에 대해 약 76.8%의 정확도를 확보하였다.

또한, 오차 행렬의 분석을 통해 모형의 정확도를 얻을 수 있다. 왼쪽 위 true negative 값과 오른쪽 아래 true positive 값이 분류의 성능을 나타낸다. Fig. 17은 오차 행렬(confusion matrix)을 보면 출력의 첫 번째 열은 실제로 정시에 출발한 항공편을 나타낸다. 해당 열의 첫 번째 행은 정시에 도착한 것으로 올바르게 예측된 항공편(true negative)을 보여주고, 두 번째 행은 지연 도착으로 예측된 항공편(false negative)을 보여준다. 출력의 두 번째 열은 실제로 지연 도착한 항공편을 나타낸다. 해당 열의 첫 번째 행은 정시에 도착한 것으로 잘못 예측된 항공편(false positive)을 보여주고, 두 번째 행은 지연 도착으로 예측된 항공편(true positive)을 보여준다.

2.3.1.2 ROC 곡선

Fig. 18은 ROC 곡선(receiver operating characteristics curve, ROC curve)은 모든 머신러닝 분류 모형에서 모형의 성능을 보여주는 그래프이다. ROC

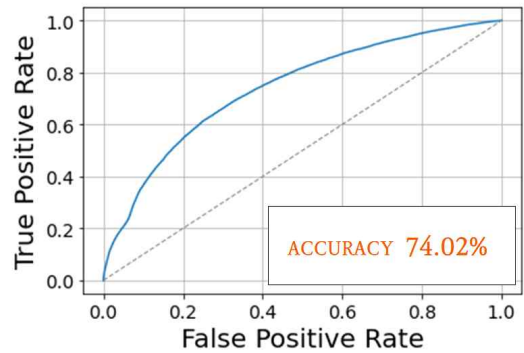


Fig. 18. ROC curve

곡선의 x-축은 false positive rate를 보여주고, y-축은 true positive rate를 보여준다. false positive rate는 실제로 정시에 도착한 항공편을 지연 도착으로 잘못 예측한 것이고 true positive rate는 실제로 지연 도착한 항공편에 대해 지연 도착한 것으로 올바르게 예측한 것이다. 그림을 보면 그래프 중앙의 점선은 올바르게 예측할 확률이 50-50임을 나타내고, ROC 곡선은 모형의 정확도를 보여준다. ROC 곡선 분석 결과 본 연구에서 제시한 모형은 약 74.02%의 정확성을 가진다.

III. 결 론

3.1 결론

본 연구에서는 2015년 1월부터 2019년 12월까지 5년간의 제주국제공항의 기상 및 운항 DATA를 기준으로 기상과 지연 간의 상관관계를 분석하고 이를 기준으로 하여 제주국제공항의 기상 현상의 양상을 심도 있게 분석하고 이를 바탕으로 출발 지연 예측 프로그램을 구축하는 것을 연구의 목적으로 하였다. 연구 결과 전체 대상 출발 운항 편수 422,218편 중 상당 부분(24.8%)인 104,593편이 지연 운항되었으며, 이 중 기상이 직접적인 원인이거나 기상 현상과 연관되어 지연된 편수는 30,350편(강수: 16,878편, 시정: 513편, 급변풍: 12,780편, 뇌전: 179편)으로 전체 출발 지연의 약 29.0%를 점하고 있는 것으로 분석되었고 전체적으로 보아 출발 지연된 편수의 상당 부분이 기상 외의 다른 원인인 항공기 연결 등의 이유로 지연이 초래되고 있는 것으로 추정, 분석되었다. 그럼에도 불구하고 계절별로 변화의 빈도와 심도가 매우 다양한 형태로 나타나는 제주국제공항의 기상과 이와 연관된 지연의 중

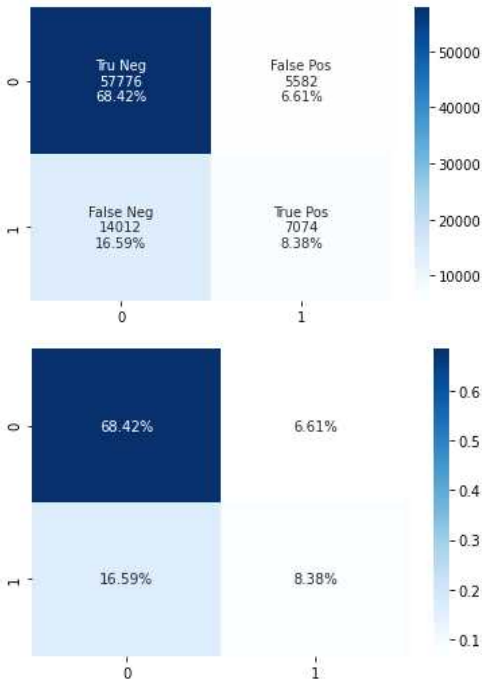


Fig. 17. Confusion matrix

요성을 근거로 누적된 기상과 운항 DATA를 최첨단의 머신러닝 기법의 하나인 랜덤 포레스트 모형을 적용하여 분석함으로써 제주국제공항의 출발 지연 예측모형을 개발하는 성과를 이루었다. 본 연구를 통해 최종 개발된 출발지연 예측모형은 개별 출발 항공편의 지연 여부를 예측할 수 있는 모형으로, 약 74% 정도의 정확성을 보이고 있다. 본 연구모형은 운항 당일 기상예보 자료를 입력자료로 활용하여 당일 제주공항 출발 항공편의 지연 여부를 예측할 수 있으며, 예측 결과는 제주공항 운영 효율성 향상에 기여할 수 있을 것으로 기대된다.

3.2 한계점 및 향후 과제

본 연구는 제주국제공항의 5년간의 기상 및 운항 DATA를 기준으로 첨단 머신러닝 기법을 이용하여 출발 편 운항 지연과 기상과의 상관관계의 분석, 그리고 출발 항공편 지연 예측 프로그램 개발에 중점을 둔 연구이다. 따라서 제주국제공항의 기상 변화에 따라 본 연구 결과를 활용함으로써 지연 예측 및 대응이 효과적으로 이루어질 것으로 기대할 수 있다. 그러나 이번 연구는 최근 5년간의 제한적인 기간 설정과 기상과 출발 항공편 지연에 국한된 연구로써 전체적인 제주국제공항의 지연의 원인을 분석하기에는 다소 아쉬움이 있었다. 제주국제공항 지연 원인의 상당 부분을 차지하는 기상 관련 DATA를 기준으로 하여 운항 DATA를 함께 분석함으로써 객관적이고 체계적으로 연구를 진행하였다는 데 큰 의미가 있으나 실질적으로 기상변수들과 출발 지연과의 뚜렷한 상관관계를 가진 큰 변수를 발견하기에는 다소 어려움이 있었다.

연구에 활용된 운항 FOIS DATA의 한계로 인하여 출발지 공항에서 이미 지연 출발 되었거나 제주국제공항에서 도착 지연으로 지연이 누적된 편과 목적지 공항 사정으로 제주공항에서 출발이 지연된 편을 구별하기가 어려워 순수한 의미의 제주출발 지연편과 기상과의 보다 정교한 분석 연구의 한계점이 있고, 제주국제공항을 기점으로 출, 도착하는 항공기의 대상 공항들의 FOIS DATA 확보를 통한 종합적인 지연 연구의 아쉬움이 있어 향후 추가적이고 기술적인 보완을 통해 제주국제공항의 지연 감소와 정시성 확보라는 정책적 대응을 위한 보다 심도 있는 연구가 수행되어야 할 것으로 생각된다. 그럼에도 불구하고 본 연구를 통하여 개별 기상변수 DATA를 분석하여 객관적인 경향성을 분석, 파악하여 시각화하였고, 향후 제주국제공항에서의

지연과 기상과의 상관관계 연구에 하나의 방향성을 제시한 것은 연구의 성과이고 또한 머신러닝이라는 첨단 기법을 사용하여 제주국제공항의 출발편 지연 예측 프로그램을 구축하였다는 것에 큰 의미가 있다고 할 수 있다.

또한, 이번 연구를 통하여 현재 제주국제공항은 최근 늘어난 운항 편수로 인하여 배분된 SLOT이 현실적으로 포화 상태이고 구조적인 항공기 연결에 의한 출, 도착 지연이 발생하고 있다는 사실을 발견할 수 있었으며, 향후 실질적인 연구를 통하여 항공사의 항공기 운항 PATTERN과 이로 인한 항공기 연결 관련 지연과 기상과의 상관관계를 종합적, 체계적인 연구의 틀에서 과학적으로 분석하여 출발지연뿐만 아니라 제주공항 출, 도착 지연의 심도 있는 원인 분석과 효율적 운영 방안 마련이 시급하다는 것이 이번 연구 과정을 통해 향후 과제로 대두되었다. 또한, 추가하여 공항 및 접근 절차의 효율성에 관한 복합적 연구의 필요성도 절실하게 대두되었다.

이번 연구에 활용한 최근 5년간의 운항 및 기상 DATA뿐만 아니라 향후 제주국제공항의 기상 및 운항 관련 BIG DATA를 활용한 체계적인 연구가 실행되어 항공기 운항과 지연 원인 간의 상관관계와 정책적 분석이 심도 있게 이루어진다면, 보다 효율적인 공항 운영을 통한 정시성 확보와 승객 불편 최소화에 기여할 것으로 기대된다.

감사의 글

본 연구는 국토교통과학기술진흥원 “데이터기반 항공교통관리 기술개발(과제번호: 21DATM- C162722-01)”의 연구지원으로 수행되었습니다.

References

1. Kim, H. Y., Yoon, J. S., Song, H. Y., and Baik, H. J., “A study on measuring airlines’ on-time performance: Focusing on departure delays”, Journal of Transport Research, 26(3), 2019, pp.1-12.
2. Ryu, Y. G., “A study on the causes of the gap between on-time departure and arrival performance in airline industry”, M.S. Thesis, Korea Aerospace University, Goyang, February

- 2009.
3. Silipo, R., and Melcher, K., "From a single decision tree to a random forest", <https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147>
 4. Raschka, S., and Mirjalili, V., "Python Machine Learning", Gilbut, 2020.
 5. Google Machine Learning Crash Course, "Classification:ROC Curve and AUC", <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>