4차 6기 세미프로젝트

## 날씨 및 대기환경이 항공기 이륙에 미치는 요인 분석

#### 2019년 12월 20일

서비스 산업 데이터를 활용한 머신러닝 분석 과정(B)

**OPTIMA** 

고영빈

윤석영

임지수

### 목 차

1. 프로젝트 개요	1
1.1 프로젝트 기획 배경 및 목표	1
1.2 구성원 및 역할	2
1.3 프로젝트 추진 일정	3
2. 프로젝트 결과	7
2.1 데이터 수집	7
2.2 데이터 분석	8
2.3 데이터 분석 결과	9
3. 기대 효과	10
3.1 향후 개선 사항	10
3.2 기대 효과	11
4. 개발 후기	12

### 프로젝트 개요 [기상/대기환경을 분석해 항공기 이륙의 최적 시나리오를 제공]

#### 1.1 프로젝트 기획 배경 및 목표

#### 프로젝트 기획 배경

- 1. 여객 항공기 수요의 증가에 따라 예측하지 못하는 상황(지연 및 결항) 또한 그 수가 빈번해지고 있다. 항공기의 지연 요인 중 승객들이 가장 많이 접하는 요인은 플랫폼 지연, 즉 A/C 지연이라고 불린다. 실제로 이 지연 상황은 해당 플랫폼의 출발 전 항공기의 기상상황으로 인한 지연 또는 항공기 정비 상황 등에 영향을 받는 요인들인데 이를 분석해서 한국 공항의 전체 지연 현황을 알아보니 기상요인에 의해 지연되는 상황이 가장 많이 발생하였다.
- 2. 항공기의 이륙은 항공사의 전적인 책임하에 "운항관리사"와 해당 항공기의 "기장" 두명의 동의로 이륙이 가능하다. 둘 중 한 사람이라도 허가하지 않으면 이륙을 하지 못하지만 현업자와 인터뷰 결과, 대부분의 결정권은 기장의 판단으로 이륙이 이루어진다. 그래서 기장의 직관적인 판단에 도움을 줄 수 있는 아이디어를 제시하고 항공사의 여객항공 운용 최적화를 위해 관련된 데이터를 수집하고 분석하여 기상요소 및 대기환경에 관련된 요인들을 알아보고 개선 방안을 제시하거나 아이디어를 찾을 수 있는 프로젝트를 진행해보고자 기획하게 되었다.

#### 프로젝트 목표

- 2 달정도 수업을 통해 배운 탐색적 데이터 분석(EDA)을 프로젝트에 활용하기 위해
  - 1. R을 활용하여 데이터 가공 후 상관분석과 시각화를 통해 분석해 본다.
  - 2. Python 을 활용하여 전처리 작업을 해보고 시각화 작업도 함께 진행하여 복습과 추가적인 학습을 목표로 한다.
  - 3. 인천공항의 기상으로 인한 항공기 지연 및 결항에 기상요인과 대기환경이 어떤 연관관계를 가지는지 알아보고 의미 있는 요소 및 환경을 분석하는 것을 목표로 한다.

#### 1.2 구성원 및 역할

이름	전공	역할	구현 부분
고영빈	산업경영공학과	팀장	프로젝트 관리
윤석영	정보경영학과	팀원	데이터 가공
임지수	사학과	팀원	자료 조사 및 해석

#### 1.3 프로젝트 추진 일정

구분	기간	활동	비고
	11/21	프로젝트 기획 및 팀 구성	첫 인사
사전 기획	11/23	PJT 주제 선정, 팀(PM/팀원) 구성	3 인/팀
	11/23	프로젝트 멘토링 [프로젝트 방향 설정 및 현업프로젝 트 소개]	현업 멘토 참여
PJT 수행 ,	11/25 ~ 11/39	프로젝트 수행	아이디어 검토 데이터 수집
<i>/</i> 완료	12/2 ~ 12/6	프로젝트 설계	데이터 전처리 일정관리

	프로젝트 멘토링 [프로젝트 점검 및 기술자문 ]	현업 멘토 참여
12/16 ~ 12/20	구현 및 테스트	Coding EDA ppt 작성
1/4 (12/20)	팀별 최종 발표 (구축 완료 보고)	최우수 한 팀 선발 멘토 평가

## 2. 프로젝트 개발 결과[항공기 지연/결항에 기상요인이 미치는 영향 분석]

2.1 데이터 수집 [인천공항 항공편 이륙 - 기상 및 대기환경 데이터]

#### Data set Title:

2016~2018 인천지역 종관기상관측 데이터 [source: 기상청]

- 측정치 비교 및 결측치 확인용

2016~2018 인천공항 기상 및 대기환경 관측 데이터 [source: 한국공항공사]

- 항공기 지연 및 결항에 미치는 기상요소 및 대기환경 변수 사용

2016~2018 인천공항 기상요인으로 인한 항공기 지연 및 결항 데이터 [source: 한국공항공사]

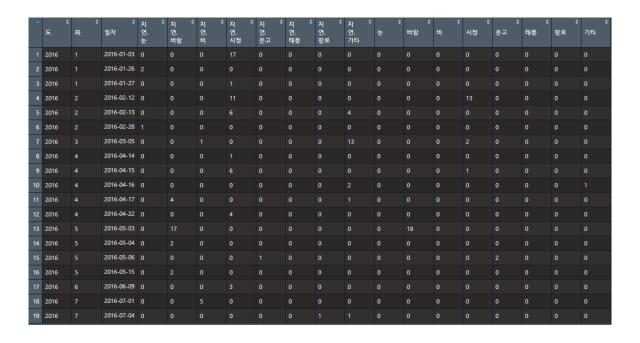
- 각 기상요인에 따른 지연 및 결항 항공편 수

2016~2018 인천지역, 서울지역 대기환경 오염도 데이터 [source: airkorea]

- 대기환경 결측치 확인 및 비교용

#### 2.2 데이터 분석 [지연/결항과 기상요인간 상관관계 분석 및 그래프 확인]

- 1. 인천공항 기상/대기환경 데이터 전처리
- → 실제 관측치에 누락이 있었는지 인천지역 종관기상관측 데이터를 확인 후 누락부분 추가 및 결측치 처리를 위해 각 기상 변수에 따라 각 데이터에 오류를 줄이기 위한 방법으로 NA처리 (구름 운량을 측정하는 octa 변수는 구름이 없는 날이 0과 NA로 나와있어 NA를 0으로 처리, 가시거리 NA는 최대시야범위로 확인돼 관측치의 최대인 10000으로 변환, 대기환경 변수들은 보간 처리)
- → 기상데이터의 이상치 처리는 종관기상관측 데이터와 비교 결과 실제 관측치로 확인

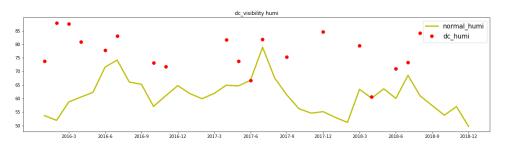


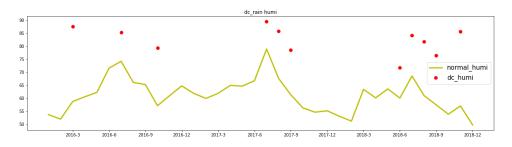
- 2. 인천공항 항공기 지연 및 결항 데이터 셋 확인 후 필요한 열 별 데이터 셋 구분
- → 기상요인 별 항공기 지연과 결항 그리고 지연+결항으로 구분
  - -> 편향된 분석을 피하기 위해 3가지로 나누어 시도
- → 각 요인 별 데이터 분석을 위해 기상데이터와 지연/결항 데이터 Set 별 Join

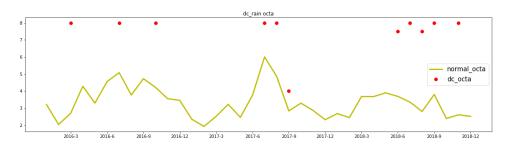
```
· str(indc1_d
'data.frame':
  year
month
day
  CO
  NO2
  PM10
PM25
SO2
  airpressure
dewpoint_temp
  humi
  sealevelpressure:
temp :
windy :
  octa
  visibility
  date
  dc_snow
dc_wind
  dc_rain
dc_visibility
  dc_cloud
  dc_typhoon
dc_fairway
dc_other
```

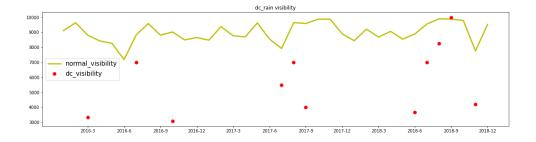
#### 3. 2016년~2018년 까지 기상요소 별 지연/결항일 때 주요 시계열 그래프 확인

#### - Python 활용

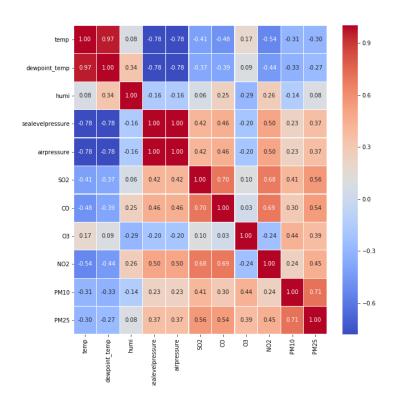








#### 4. 기상요인과 대기환경요인 각각의 상관관계 확인



#### 2.3 데이터 분석 결과

항공기지연일때 기상:	요인간상	관관계						
	d_snow	d_wind	d_rain	d_visibility	#d_cloud	#d_typhoon	d_fairway	d_other
airpressure	0.23330	0.25676	-0.28225	0.25541	NA			
dewpoint_temp	0.33305		0.23845		NA		0.11402	0.16411
humi	0.55387	-0.12065		0.24269	NA		0.13047	0.03784
sealevelpressure	0.23267	0.25676	-0.28225	0.25536	NA			-0.04872
temp	0.05884		0.30826		NA			0.19217
windy	-0.12426	0.05031	0.28300		NA		0.04874	0.07293
octa	0.40380	0.28448	0.18283	0.08508	NA	NA	0.25242	0.16738
visibility		-0.20276	0.06501	-0.31982	NA	NA		

항공기가 각 기상요인 별 지연일 때 눈 -> 습도, 운량, 이슬점온도, 가시거리 바람 -> 기압 비 -> 바람 시정 -> 가시거리, 온도 항로 -> X

# 항공기지연일 때대기오염간상관관계 d\_snow d\_wind d\_rain d\_visibility #d\_cloud #d\_typhoon d\_fairway d\_other CO 0.29847 0.22546 -0.474058 0.4173 NA 1 -0.079461 -0.465285 NO2 0.35126 0.78181 -0.265159 0.3000 NA -1 -0.080929 -0.173079 O3 -0.45061 -0.30757 -0.121572 -0.2368 NA 1 -0.104226 -0.153192 PM10 0.03789 0.07687 -0.001369 0.1686 NA -1 -0.003915 0.026898 PM25 0.28416 0.11648 -0.079629 0.2293 NA -1 -0.074054 -0.001804 SO2 0.29374 0.28904 -0.332773 0.2130 NA -1 -0.007728 -0.521821

항공기가 각 기상요인 별 지연일 때 눈 -> NO2, CO, SO 시정 -> CO

항 <del>공</del> 편지연 및 결항일 [	대기상요(	인간 상관	관계					
	dc_snow	dc_wind	dc_rain	dc_visibility	#dc_cloud#dc	_typhoon	dc_fairway	dc_other
airpressure	0.2468	0.07257		0.26541	-0.11457			-0.04928
dewpoint_temp	0.3217		0.25686				0.11402	0.16289
humi	0.5524	-0.05421		0.20623	0.01056	NA	0.13047	0.47071
sealevelpressure	0.1055	0.07220		0.37831	0.08719	NA	0.06101	0.17843
temp	0.0451	-0.30082	0.31915				0.06959	0.19160
windy		0.16480	0.20116				0.04874	0.07123
octa	0.4034	0.34146	0.18824	0.05703	NA	NA	0.25242	0.16953
visibility			0.07329		0.83138	NA		-0.08824

항공편 기상요인 별 지연+결항일 때

는 -> 습도, 이슬점온도, 운량 바람 -> 운량, 온도 비 -> 온도 구름 -> 운량

# 항공편지연및결항일 때 대기오염간 상관관계 dc\_snow dc\_wind dc\_rain dc\_visibility #dc\_cloud #dc\_typhoon dc\_fairway dc\_other C0 0.30894 0.11513 -0.4765 0.4157 -0.6109 1 -0.079461 -0.462309 N02 0.37156 0.67037 -0.2204 0.3203 0.5475 -1 -0.080929 -0.162092 O3 -0.46039 -0.26870 -0.1559 -0.2285 0.5016 1 -0.104226 -0.151760 PM10 0.04203 0.03425 -0.0294 0.1820 0.4932 -1 -0.003915 0.033220 PM25 0.28934 0.04989 -0.1040 0.2363 0.3662 -1 -0.074054 0.005136 S02 0.30813 0.17838 -0.3463 0.2188 -0.4763 -1 -0.007728 -0.515853

항공편 기상요인 별 지연+결항일 때 눈 -> CO, NO2 가시거리 -> CO

#### 결과 해석

해당 상관관계 결과를 분석 하였을 때, 가시거리에 중요한 영향을 미치는 변수가 일반적으로 생각할 수 있는 기상요인들 보다 대기 중 CO, NO2 물질의 농도가 많은 관계가 있다는 것을 발견하였다. 그리고 실제 기장이 각 기상요인 별 지연/결항을 결정할 때 당시 기상상황이 직접적인 관련사항이라기 보다 여러 상황을 종합해서 판단을 내린 것으로 알게 되었고 전체적으로 구름의 양과 가시거리, 이슬점 온도가 상대적으로 많은 지연/결항에 영향을 미친다는 것을 분석할 수 있었다

#### 3. 기대 효과

#### [항공사 운용에 도움이 되는 기상/대기환경 분석 결과 제공]

3.1 향후 개선 사항 [데이터 세분화에 대한 아쉬움과 분석 의의 찾기]

본 프로젝트에서 보다 의미 있는 결과를 도출하지 못한 것은 데이터 부분에서 아쉬움이 있었기 때문이라 생각한다. 우선, 인천 공항의 지연 및 결항 데이터 셋의 구조가 너무 크게 잡혀있었다. 단순하게 기상으로 인한 지연과 결항의 데이터만이 있으며, 지연 시간을 비롯하여 세부적인 지연 및 결항의 외부요인에 대한 데이터는 생략되어 있었다. 마찬가지로 항공기의 기종, 목적지, 이용활주로(활주 방위), 이륙 시간 등 항공편에 대한 정보가 조금 더 주어졌다면, (혹은 제공되는 자료가 존재하였다면) 항공기에 따른 지연 및 결항에 대한 분석이 더 용이하였을 것이며 활용방안 또한 보다 다양하였을 것이라고 본다. 만약, 해무 / 연직시어(상층과 하층의 풍속 차이)에 대한 데이터가 있다면, 사고와 관련된 데이터를 종합하여 항공 안전에 까지 도움을 줄 수 있을 것이라 생각한다.

예시를 하나 들자면, 분석 과정에서 일산화탄소와 지연 및 결항의 상관관계를 확인하였다. 이에 항공기가 지연 및 결항 과정에서 계속하여 비행 엔진을 켜둠에 따라 일산화탄소의 배출량이 증가하였기 때문이라 생각할 수도 있었는데, 이를 구체적인 데이터로 확인하고자 하였다. 하지만 항공기가 배출하는 가스(일산화탄소를 비롯한 다수의 배출가스)에 대한 데이터는 정확하게 측정되지 않고 있었다. 이른바 LTO(Landing and Take-off), 이륙, 착륙 시에 발생하는 배출가스의 양이 어느 정도인지 항공기 개별로 측정된 데이터가 없었으며, 이를 세분화하여 측정한 데이터 또한 없었다. 만약 이 데이터가 존재하였다면, 항공기 지연 및 결항 시간에 따른 배출량 증가의관계가 있음을 파악할 수 있었다고 본다.

추가적으로, 본 분석의 목적이 기장 및 운항 관리사의 판단을 돕기 위한 것인 만큼, 데이터 분석결과로 제안할 결과물이 필요로 되며, 머신러닝을 통한 예측이 가능해진다면 기장과 운항관리사의 주관적인 판단이 아닌 이전의 과거 데이터와 현재 기상 상황의 데이터를 통해 보다효율적으로 항공기의 운항을 결정할 수 있게 될 것이다.

#### 3.2 기대 효과 [본문 : 10pt / 맑은 고딕]

프로젝트를 통해 가시거리에 중요한 영향을 미치는 변수가 일반적으로 생각할 수 있는 기상요인들 보다 대기 중 CO, NO2 물질의 농도가 많은 관계가 있다는 것을 발견하였으며, 이에 대한 데이터를 기장 및 운항관리사에게 전달함으로써 보다 효율적인 지연 및 결항 결정이 가능할 것이라고 본다. 특히, 항공기 이착륙 사고에서 이륙 3분, 착륙 8분의 시간이 항공 사고 발생의 가장 높은 비율을 차지하고 있는 만큼, 이착륙에 영향을 미칠 수 있는 데이터를 보다 다양하게 수집하고, 그 연관성을 파악한다면 사고 예방 데이터로써 활용될 수 있다. 즉, 항공 운항에 영향을 주는 데이터들을 통해 지연 및 결항을 결정할 뿐 아니라, 항공 안전, 기후에 따른 최적의 항로 제공 등으로 활용할 수 있으며, 발생하는 금전적인 손해 또한 최소한으로 감소시킬 수 있고 항공 승객 또한 보다 안전한 운항 서비스를 제공받을 수 있게 된다.

#### 4. 개발 후기



성명	후기
고영빈	이번 프로젝트를 진행하면서 많은 사람들을 만나게 되고 많은 데이터를 접하게 되면서 식견이 많이 생기고 산업 전반에 대한 기초지식이 많이 필요한 것을 깨닫게 되었습니다. 앞으로 공모전이나 프로젝트를 진행할 때 결과물을 미리 생각해보고 데이터를 활용한 개선 결과물이나 제안 내용들도 필수적으로 생각해 분석에 의의를 강조하고 활용할 수 있는 프로젝트를 만들 수 있도록 노력하겠습니다. 참 좋은 경험을 만들어준 프로젝트였습니다. 감사합니다.

	T
윤석영	이번 프로젝트를 진행하면서 해당 데이터들의 이해와 시각화가 많이 중요하다는 것을 느꼈습니다. 정상적으로 운행하였을 때의 데이터와 지연이나 결항이 되었을 때의 데이터 비교를 숫자로만 보다가 나중에 그래프로 그려보았을 때 막연하게 느껴지던 부분들이 명확하게 다가왔습니다. 이로 인하여 시각화까지 수행한 후에 분석에 들어가야 한다는 것을 알게 되었습니다. 다음에 프로젝트를 수행할 때에는 좀 더 이해한 후에 진행하여 더 나은 결과를 보이도록 하겠습니다
임지수	본 세미 프로젝트는 지금까지 배운 것과 앞으로의 학습 방향성에 대해 생각하게끔 하는 좋은 기회였습니다. 약 2 달 동안 배운 R, Python 을 비롯한 기본적인 데이터 분석 도구를 이용하여 데이터 수집 및 전처리 과정을 수행함으로써 부족한 부분을 확인하였고, 아직 부족한 부분이 있음 또한 알게 되었습니다. 이론과 기본적인 연습문제를 통해서만 학습하는 것이 실제 데이터 분석 과정과 상당히 다르다는 점을 깨달았으며, 앞으로는 공모전을 통해 데이터를 실용적인 방향으로 가공 및 분석하는 과정을 조금 더 연습해보고자 합니다. 프로젝트를 진행하는 과정 내에서 팀원이 본 과정을 나가는 경우와 같이 예상치 못한 상황을 직면하기도 하였지만, 반대로 서로 다른 의견을 가진 팀원들과 하나의 주제를 정하고 각자의 역할을 나눠 실행하는 과정에서 팀 프로젝트의 의의를 깨닫기도 하였습니다. 아직은 개인적으로도 많이 아쉬운 부분이 있지만, 이를 계기로 하여 다음의 공모전 및 파이널 프로젝트는 보다 완성도 있는 결과를 보이겠습니다.