



항공지연예측

판다곰 - 박성은, 신태범

목차



문제정의



데이터 분석



데이터 전처리

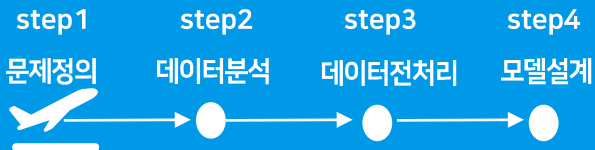


모델설계

step1

문제정의

어떻게 지연을 예측할 것 인가?



지연예측 효용



- 여행대기시간 감소
- 공항 혼잡방지



- 공항 자원 운영
효율화



정시성 확보

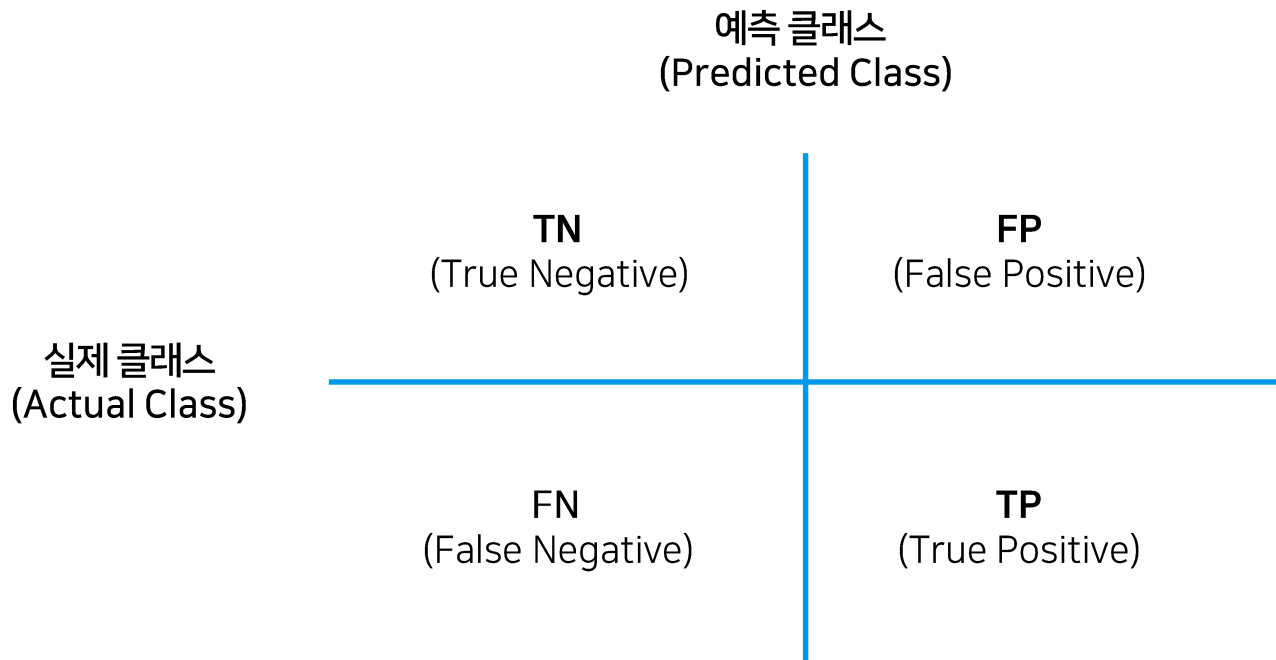


- 효율적 시간배분
- 운항편수추가



- 대기시간 감소
- 연료절감

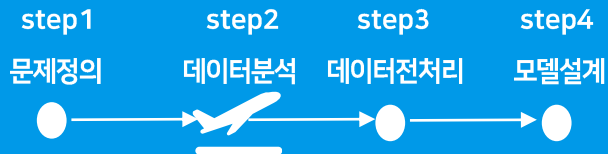
METRIX 설정



step2

데이터분석

데이터들은 어떤 특성을
가지고 있는가?

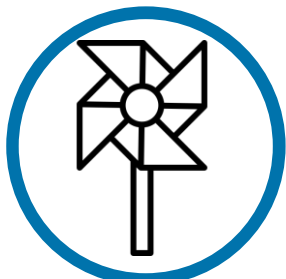


항공운항데이터

- AFSNT 운항 실적
- AFSNT_DLY 지연예측 TEST SET

항공기상데이터

- 1시간 단위로 기록된 **METAR 데이터** 이용
(출처: 항공기상청 공항통계자료)
- 데이터 내의 자연과 관련 높은 변수 사용



WD
풍향



RN
강수량



VIS
시정



CA_TOT
전운량



WC
일기

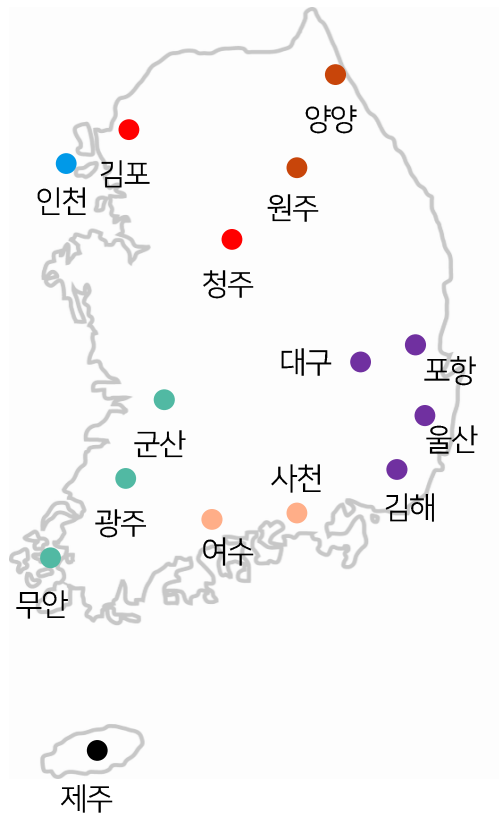
항공기상데이터

● METAR 데이터 : 1시간 단위

TM	WD	VIS	WC	RN	CA_TOT
2016123124	9	600	NULL	NULL	NULL
2017010101	10	600	10	NULL	6
2017010102	9	600	10	NULL	6
2017010103	9	290	NULL	NULL	NULL
2017010104	10	320	NULL	NULL	NULL
2017010105	9	230	NULL	NULL	NULL
2017010106	10	280	10	NULL	6
2017010107	10	350	10	NULL	3
2017010108	13	420	10	NULL	6
2017010109	9	420	10	NULL	7

NULL 데이터는 기록이 없다는 것?
전처리시 0으로 변경

항공기상데이터



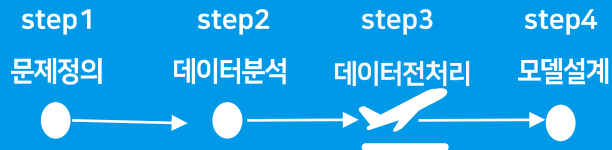
● 모든 공항에 대한 기상 누적 데이터가 존재하지 않음.

➡ 데이터가 존재하는 기준으로 직선거리 가까운 공항끼리 기상데이터 공유

step3

데이터전처리

데이터들을 어떻게
가공하였는가?



데이터 전처리

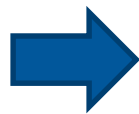
- 범주형 변수

Sklearn.LabelEncoder로 수치화

- Schedule 컬럼 생성

운행데이터 SDT_YY, SDT_MM, SDT_DD, STT 시간 이용

SDT_YY	SDT_MM	SDT_DD	STT
2017	01	01	10:05



Schedule
2017-01-01-10:05

이 Schedule 컬럼 이용 해 운행데이터와 기상데이터 병합

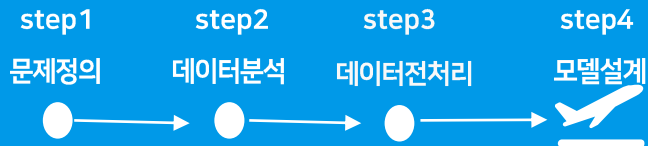
새로운 열생성

- 각 날짜를 기준으로 각 컬럼을 mean처리 (`data.groupby(date).mean()`)
- **Delay**
해당 날짜의 DLY(지연)의 평균이 0.3 이상이면 1, 아니면 0
- **Probability**
각 날짜의 기상 데이터를 고려한 지연 확률 - `xgbclassifier.predict_proba()` 이용
- **DIFF**
STT(계획된 출발 시각)와 ATT(실제 출발 시각)의 차이
- **DIFF_predict**
다른 기상 컬럼을 이용해 예측한 DIFF- `xgbregressor` 이용

step4

모델설계

문제해결을 위해 어떤 모델을
만들었는가?



- **A Deep Learning Approach to Flight Delay Prediction** (Kim et al.)

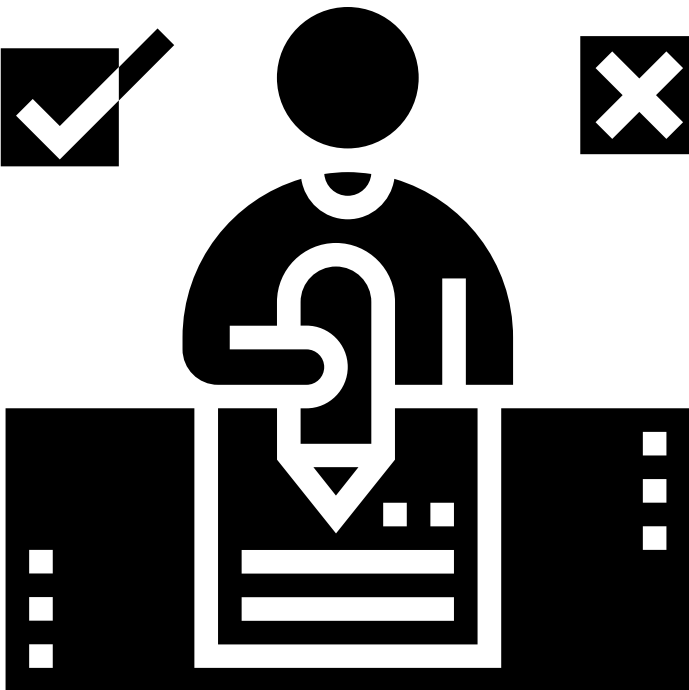
Published in IEEE DASC(Digital Avionics Systems)

선행 연구

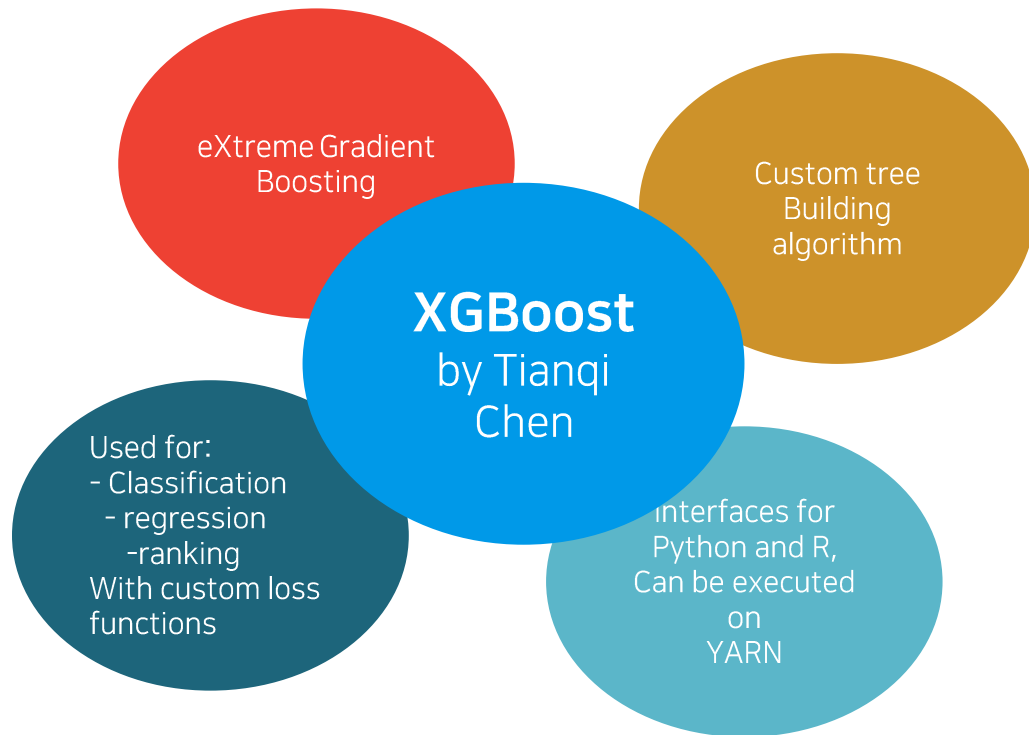
Day-Delay-Status를
비행기 지연여부에측에 사용



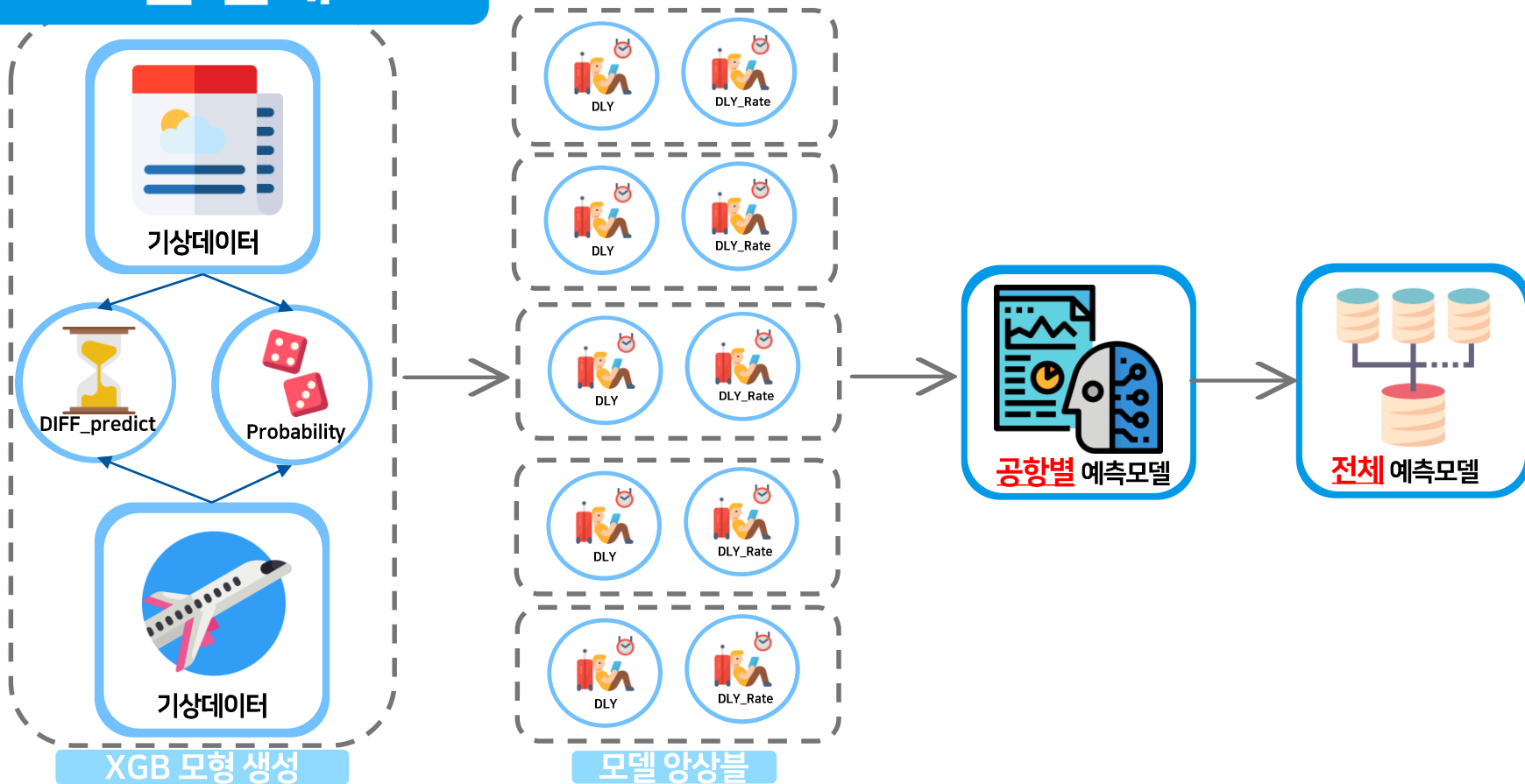
하루가 independent 하다고
가정



모델 설계



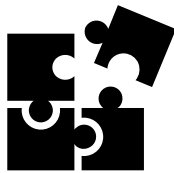
모델 설계



모델 설계 절차



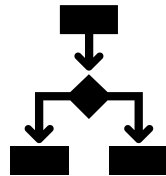
Step1
Probability 생성
DIFF예측



Step2
XGB모형 생성

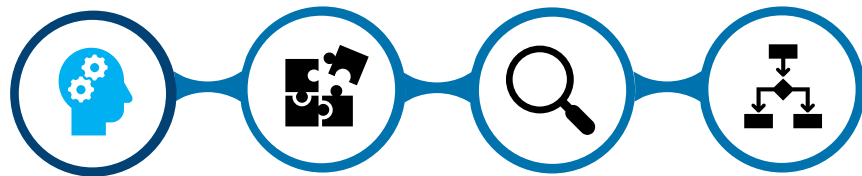


Step3
Hyperparameter
Search



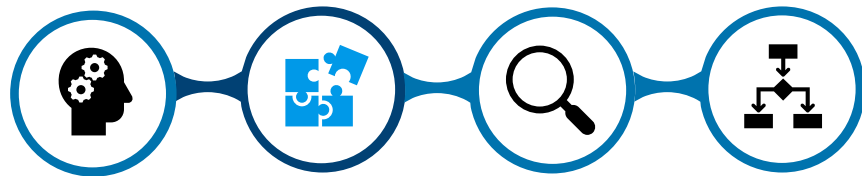
Step4
XGB 모형 앙상블
결과출력

STEP1



- 날짜별 지연 Probability 구하기 By XgbClassifier
- DIFF를 날짜별로 예측하기 By XgbRegressor

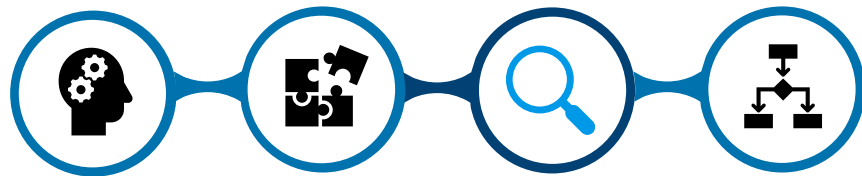
STEP2



- 각 편명에 제공되는 정보들을 이용하여 xgb 모델을 만든다.

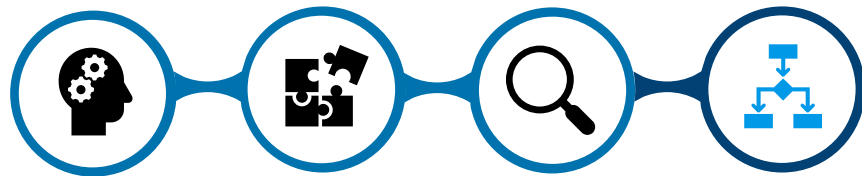
(정보 : AFSNT_DLY 에 있는 컬럼 + 2018년 09월 기상 정보)

STEP3



- 공항 별 최적 **hyperparameter search** with f1_score

STEP4



- **앙상블**

총 5개의 xgb모형을 만들고 결과치를 평균 내서 최적모델 설계.

- 최적모델로 지연여부, 지연률 도출

- AFSNT_DLY의 범주형 변수 원상 복구 with `inverse_transform()` function