중간평가 이후 첫 DAB 회의

(교통량 이상탐지 -〉 영향이 있는 공연, 없는 공연의 분류모델 제작
분류모델- 종속변수: 영향력 여부, 독립변수: 공연정보/ 중요한 feature? 공연에 온 사람 규모)

1. 지금까지 내용 정리

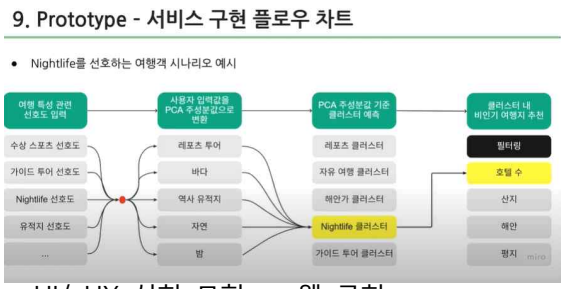
1) 예측인원 행사모형
- 인터파크 공연 데이터를 웹크롤링해서 이를 바탕으로 공연정보를 벡터화할 수 있는 모델을 구축함.
- 해당 모델을 통해 올림픽공원 데이터 벡터화 진행 (벡터화 이유: 유사공연을 용이하게 찾기 위함.)
- 올림픽공원 데이터 (vector3(index,n): 올림픽공원 데이터 중 해당 index 공연과 유사한 공연 n개를 도출해주는 함수)

변수 수정으로 예측률 상승 평균 4%, 회귀사용 14%, ANN 18%

2) 교통량 이상치 탐지모형
- 교통량 및 대중교통 데이터를 이용해 행사들이 교통량에 미친 영향을 파악하는 모델.
- 올림픽공원 입차, 출차 데이터 〉 도로 지점 1시간 단위 교통량 〉 지하철 역별 1시간 단위 승하차 인원 기록 데이터
- (표준화 잔차 2 이상인 값들이 이상치)- 데이터 프레임과 표준화 잔차를 입력하면 해당하는 행사를 행사 데이터에서 찾아주는 탐지 함수 생성

2. 앞으로의 논의
1) 교통량 이상탐지 모델과 행사인원 예측모델 간의 연관성

2) 프로젝트 사업화
- 비즈니스 모델 제안
- 사업화 제안 플로우차트



- UI/ UX 실현 모형 or 웹 구현

3) 최종발표 ppt
- 목차논의

- 팀원소개, 제안 배경 및 필요성, 프로젝트 목표 등 (지금 만들어 놓을 수 있는 것들은 미리 만들어 놓는 게 편할 듯.)
- 분석과정 전체 Overview
(처음 듣는 사람들이 우리 내용 알아들을 수 있도록)
(최종 때는 전처리는 제외하고 성과 위주로)
양식 노션에 올려두었으니 참고

타 연구 비교- 행사인원 예측모델

1. 급식소 식수 예측 모델
1) 행사데이터 변수 추가

그림 3-1 식수 예측 모델 개발 연구 모형

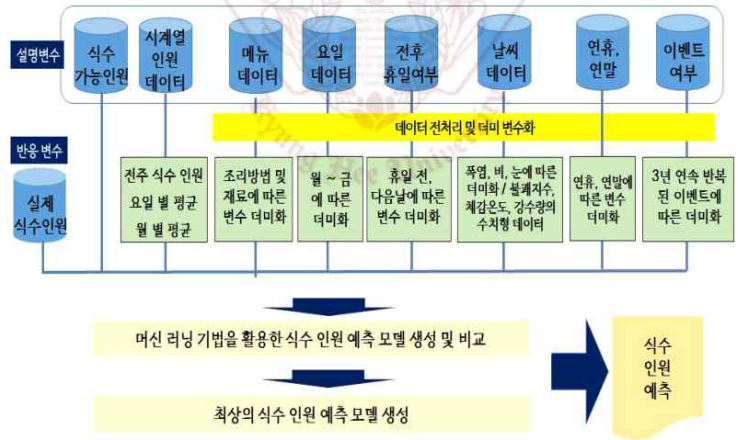
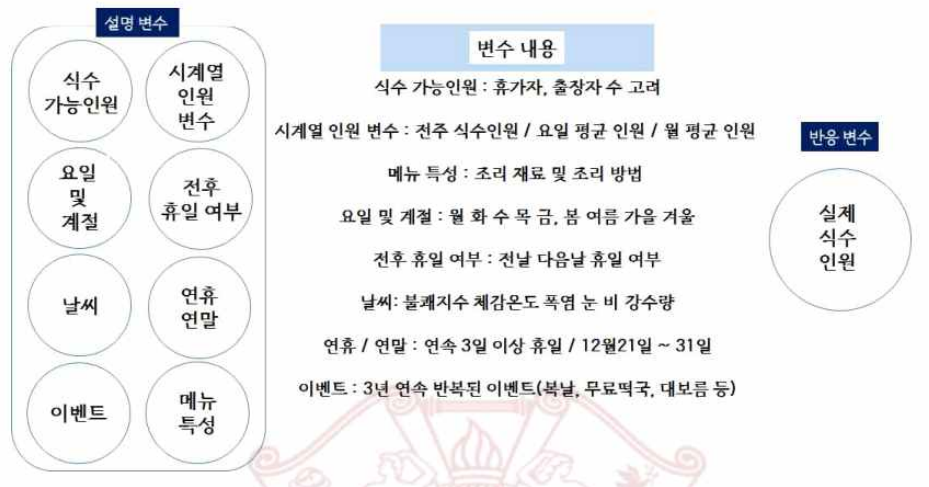


그림 3-3 식수 예측 모델링에 사용한 변수 요약



날씨 데이터 뿐만 아니라 변수 추가하는 것도 방법.

추가 변수

요일 - 월~일에 따른 더미화

계절 변수 - 봄~겨울에 따른 더미화

전후 휴일여부 - 휴일 전후에 따른 변수 더미화

연휴, 연말 여부 - 연휴, 연말에 따른 변수 더미화 (3일 연속 연휴)/ (12/21~31)

날씨 데이터- 폭염, 비, 눈에 따른 더미화/ 체감온도, 강수량의 수치형 데이터

+. (요일별 평균인원/ 월별 평균인원/ 전주 인원)=> 회귀분석으로 해당 변수의 설명력을 파악한 뒤 변수 포함여부 결정

표 3-8 요일 별 식수 (요일 평균) 인원 생성표

대상일	요일	요일 평균
20180420	월	1336
20180421	화	1208
20180422	수	1212
20180423	목	1151
20180424	금	1001
----	---	----

→ 요일 평균 인원 변수 추가 예시

2) 날씨 데이터

날씨 데이터

측정일자 (MESURE_DE)

측정요일명 (MESURE_WKDAY_NM) – 월요일 (1) ~ 일요일 (7)로 코딩 필요

주말여부 (WKEND_AT)

측정온도값 (MESURE_TP_VALUE)

강수형태코드 (PRCPT_STLE_CD)- 없음 (0), 비 (1), 비/눈 (2), 눈 (3)

-) 이정도 변수들만 살리는 게 좋을 듯.

강수량 포함 여부는 논의

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	MESURE_D	MESURE_W	WKEND_A	MESURE_H	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	CRRSPND_	MESURE_T	PRCPT_STL	PRCPT_STL	TIME_ISE_H	MESURE_H	MESURE_F	MESURE_W	WD_FLAG_
2	20120107	토요일	Y	0	0	0	0	0	0	0	0	0	0	-8.3	0	없음	0	71	0.5	0	N
3	20120107	토요일	Y	1	0	0	0	0	0	0	0	0	0	-5.8	0	없음	0	62	0.5	0	N
4	20120107	토요일	Y	2	0	0	0	0	0	0	0	0	0	-3.7	0	없음	0	53	0.9	0	N
5	20120107	토요일	Y	3	0	0	0	0	0	0	0	0	0	-1.2	0	없음	0	45	0.9	0	N
6	20120107	토요일	Y	4	0	0	0	0	0	0	0	0	0	0.3	0	없음	0	37	0.5	0	N
7	20120107	토요일	Y	5	0	0	0	0	0	0	0	0	0	1	0	없음	0	34	0.7	0	N
8	20120107	토요일	Y	6	0	0	0	0	0	0	0	0	0	1.5	0	없음	0	30	1.3	0	N
9	20120107	토요일	Y	7	0	0	0	0	0	0	0	0	0	1.3	0	없음	0	32	0.9	0	N
10	20120107	토요일	Y	8	0	0	0	0	0	0	0	0	0	0.2	0	없음	0	39	0.4	0	N
11	20120107	토요일	Y	9	0	0	0	0	0	0	0	0	0	-1.1	0	없음	0	45	0.9	0	N
12	20120107	토요일	Y	10	16	0	0	0	0	0	0	0	0	-1.9	0	없음	0	50	0.4	0	N
13	20120107	토요일	Y	11	31	2	0	0	0	2	0	0	0	-3	0	없음	0	56	0.1	0	N
14	20120107	토요일	Y	12	25	11	0	0	0	11	0	0	0	-3.8	0	없음	0	60	0.4	0	N
15	20120107	토요일	Y	13	33	8	0	0	0	8	0	0	0	-4.5	0	없음	0	59	0.6	0	N
16	20120107	토요일	Y	14	7	14	0	0	0	14	0	0	0	-4.9	0	없음	0	59	0.3	0	N
17	20120107	토요일	Y	15	14	18	0	0	0	18	0	0	0	-5.5	0	없음	0	69	0.4	0	N
18	20120107	토요일	Y	16	9	19	0	0	0	19	0	0	0	-6	0	없음	0	68	0.2	0	N
19	20120107	토요일	Y	17	2	10	0	0	0	10	0	0	0	-6.4	0	없음	0	72	0.3	0	N
20	20120107	토요일	Y	18	3	40	0	0	0	40	0	0	0	-6.8	0	없음	0	69	0.6	0	N
21	20120107	토요일	Y	19	0	9	0	0	0	9	0	0	0	-6.8	0	없음	0	73	0.6	0	N
22	20120107	토요일	Y	20	1	1	0	0	0	1	0	0	0	-6.3	0	없음	0	67	0.5	0	N

시간대별 데이터이기에 일별 데이터로 바꾸고 변수 일부 수정이 필요

날짜별 날씨데이터

- 같은 날짜별로 그룹 만들어주고 최고기온, 최저기온, 비, 눈 컬럼 생성
- 더위 체감 지수 높음 (28℃ 이상) & 한파주의보 (-12℃ 이하)로 이상기온 여부 컬럼 생성

최종 날씨 데이터 컬럼

날짜	요일 (더미화)	주말여부	최고기온	최저기온	이상기온 여부	비	눈

공연이 하루에 걸쳐 진행되지 않는 경우가 더 많은 데 이때 날짜는 어떤 기준으로...?

---

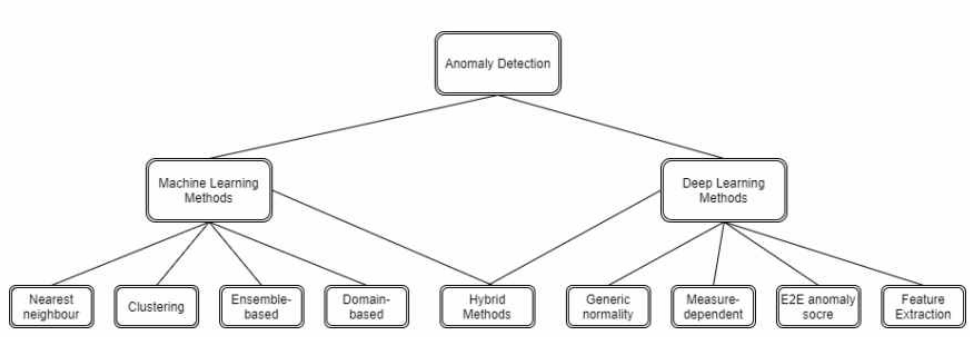
3) 다른 적용 기법 (다중선형회귀, ANN 이외)
- LASSO or Ridge
- Bagging
- Boosting
- Random Forest

→ 오차범위 성능지표 활용해 모델 평가

행사인원 참여인원 역할분담
- 벡터화 모델
- 날씨데이터 리서치 ex. 불쾌지수
- 날씨데이터 컬럼 생성
- 행사데이터 컬럼 생성
- 방법론 여러 개 적용

타 연구 비교- 이상치탐지모델

1. 오토인코더를 사용한 이상탐지 모델의 비교분석 및 이상치 판별 기준 제안



(딥러닝)

1) Autoencoder

: 오토인코더의 성능향상
- DAE (잡음 제거)
- Sparse AE
- VAE

2) GAN
- AnoGAN
- GANomaly

3) LSTM

RNN	Stacked LSTM LSTM and GRU Hybrid LSTM+OC–SVM/SVDD
CNN	FCN ConvLSTM
AutoEncoder	LSTM–ED MSCRED Multi–modal DAE ConvLSTM–AE
Generative Models	GAN Variational Inference

→ 모델의 성능 평가의 척도로는 정확도(Accuracy), AUC ROC, 조화평균(F1 Score), AUC PRC