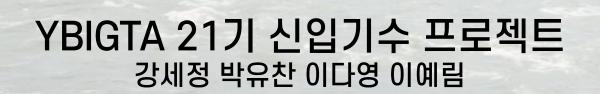
## 한강 수위 예측 Al

CNN-LSTM을 활용한 한강 수위 예측



## **목차**

1 Motivation

5 Preprocessing

2 Data

6 Experiments

3 Process

7 Improvement

4 Model

# Motivation

#### 1. Motivation

- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### 팔당댐 홍수 안전운영에 따른 한강 수위예측 AI 경진대회

알고리즘 | 정형 | 회귀 | 시계열 | RMSE

- ₩ 상금:총1,100만원
- () 2022.07.25 ~ 2022.09.02 17:59 (+ Google Calendar

**ஃ** 1,161명 📋 D-7



#### 대회 배경 및 목표

- 팔당댐의 홍수 안전운영에 따른 서울시내 한강 주요지점의 수위를 예측
  - → 홍수재해로 인한 피해를 미연에 방지하고 최소화 할 수 있을 것으로 기대
- 목표 : 팔당댐 방류에 따른 <u>서울시내 한강 주요다리 수위예측</u>

#### 평가

- 기준: RMSE / R\_Squared\_Score



- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

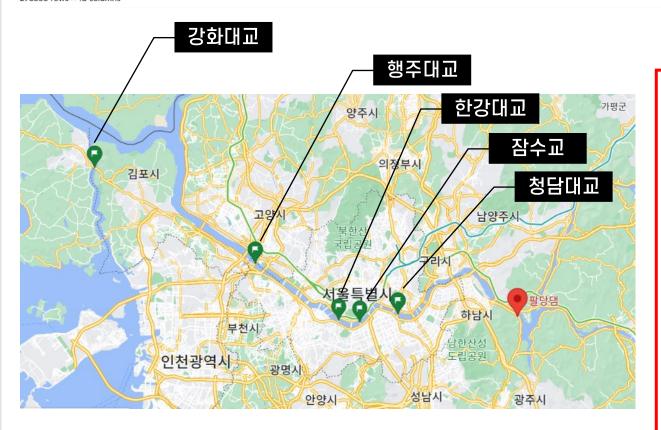
	날짜	현재수위	유입량	저수량	공용량	방류량	강화대교_조위	청담대교_유량	잠수교_유량	한강대교_유량	행주대교_유량	대곡교_강수량	진관교_강수량	송정동_강수량	청담대교_수위	잠수교_수위	한강대교_수위	행주대교_수위
0	2012-05-01 00:00	24.800	555.00	219.07	24.93	555.00	445.0	469.05	0.0	729.80	540.18	0.0	0.0	0.0	310.7	300.2	290.0	275.3
1	2012-05-01 00:10	24.794	464.60	218.86	25.15	562.90	449.0	498.00	0.0	731.48	540.18	0.0	0.0	0.0	314.7	300.2	290.0	275.3
2	2012-05-01 00:20	24.789	478.10	218.69	25.31	576.40	451.0	490.68	0.0	726.42	540.18	0.0	0.0	0.0	313.7	301.2	290.0	275.3
3	2012-05-01 00:30	24.789	464.80	218.69	25.31	563.10	452.0	476.21	0.0	726.42	552.17	0.0	0.0	0.0	311.7	301.2	290.0	276.3
4	2012-05-01 00:40	24.789	478.10	218.69	25.31	576.40	450.0	476.21	0.0	707.17	564.29	0.0	0.0	0.0	311.7	301.2	291.0	277.3
276331	2022-07-18 23:10	25.040	259.23	212.86	31.14	259.23	510.0	319.84	NaN	-456.41	974.40	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276332	2022-07-18 23:20	25.040	260.46	212.86	31.14	260.46	492.0	314.01	NaN	-717.30	1006.88	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276333	2022-07-18 23:30	25.040	259.37	212.86	31.14	259.37	475.0	387.55	NaN	-843.37	1039.90	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276334	2022-07-18 23:40	25.040	259.13	212,86	31.14	259.13	458.0	454.91	NaN	-1023.37	1073.46	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276335	2022-07-18 23:50	25.040	258.16	212,86	31.14	258.16	442.0	582.21	NaN	-1049.44	1090.45	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276336 rov	ws × 18 columns																	

팔당댐 데이터 한강 다리 유량 데이터 강수량 데이터

현재수위	유입량	저수량	공용량	방류량	강화대교_조위	청담대교_유량	잠수교_유량	한강대교_유량	행주대교_유량	대곡교_강수량	진관교_강수량	송정동_강수량
24.80	555.00	219.07	24.93	555.00	445.0	469.05	0.0	729.80	540.18	0.0	0.0	0.0
24.79	464.60	218.86	25.15	562.90	449.0	498.00	0.0	731.48	540.18	0.0	0.0	0.0
24.78	478.10	218.69	25.31	576.40	451.0	490.68	0.0	726.42	540.18	0.0	0.0	0.0
24.78	464.80	218.69	25.31	563.10	452.0	476.21	0.0	726.42	552.17	0.0	0.0	0.0
24.78	478.10	218.69	25.31	576.40	450.0	476.21	0.0	707.17	564.29	0.0	0.0	0.0
25.04	259.23	212.86	31.14	259.23	510.0	319.84	NaN	-456.41	974.40	0.0	0.0	0.0
25.04	260.46	212.86	31.14	260.46	492.0	314.01	NaN	-717.30	1006.88	0.0	0.0	0.0
25.04	259.37	212.86	31.14	259.37	475.0	387.55	NaN	-843.37	1039.90	0.0	0.0	0.0
25.04	259.13	212.86	31.14	259.13	458.0	454.91	NaN	-1023.37	1073.46	0.0	0.0	0.0
25.04	258.16	212.86	31.14	258.16	442.0	582.21	NaN	-1049.44	1090.45	0.0	0.0	0.0

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

	날짜	현재수위	유입량	저수량	공용량	방류량	강화대교_조위	청담대교_유량	잠수교_유량	한강대교_유량	행주대교_유량	대곡교_강수량	진관교_강수량	송정동_강수량	청담대교_수위	잠수교_수위	한강대교_수위	행주대교_수위
0	2012-05-01 00:00	24.800	555.00	219.07	24.93	555.00	445.0	469.05	0.0	729.80	540.18	0.0	0.0	0.0	310.7	300.2	290.0	275.3
1	2012-05-01 00:10	24.794	464.60	218.86	25.15	562.90	449.0	498.00	0.0	731.48	540.18	0.0	0.0	0.0	314.7	300.2	290.0	275.3
2	2012-05-01 00:20	24.789	478.10	218.69	25.31	576.40	451.0	490.68	0.0	726.42	540.18	0.0	0.0	0.0	313.7	301.2	290.0	275.3
3	2012-05-01 00:30	24.789	464.80	218.69	25.31	563.10	452.0	476.21	0.0	726.42	552.17	0.0	0.0	0.0	311.7	301.2	290.0	276.3
4	2012-05-01 00:40	24.789	478.10	218.69	25.31	576.40	450.0	476.21	0.0	707.17	564.29	0.0	0.0	0.0	311.7	301.2	291.0	277.3
276331	2022-07-18 23:10	25.040	259.23	212.86	31.14	259.23	510.0	319.84	NaN	-456.41	974.40	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276332	2022-07-18 23:20	25.040	260.46	212.86	31.14	260.46	492.0	314.01	NaN	-717.30	1006.88	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276333	2022-07-18 23:30	25.040	259.37	212.86	31.14	259.37	475.0	387.55	NaN	-843.37	1039.90	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276334	2022-07-18 23:40	25.040	259.13	212,86	31.14	259.13	458.0	454.91	NaN	-1023.37	1073.46	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276335	2022-07-18 23:50	25.040	258.16	212.86	31.14	258.16	442.0	582.21	NaN	-1049.44	1090.45	0.0	0.0	0.0	0.0	0.0	0.0	0.0
276336 roy	vs × 18 columns																	

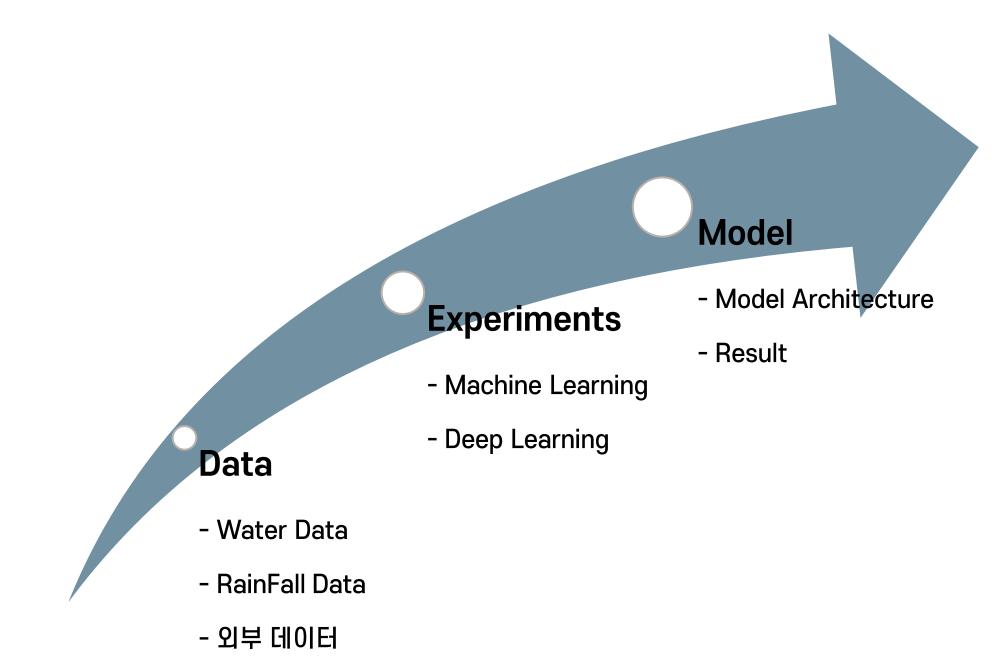


#### 한강 다리 수위 데이터

청담대교_수위	잠수교_수위	한강대교_수위	행주대교_수위
310.7	300.2	290.0	275.3
314.7	300.2	290.0	275.3
313.7	301.2	290.0	275.3
311.7	301.2	290.0	276.3
311.7	301.2	291.0	277.3
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0



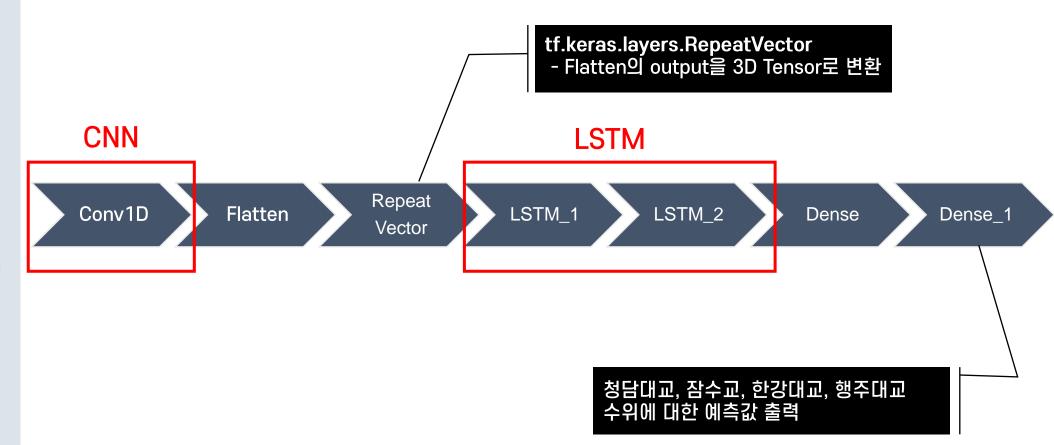
- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement





## - **† TensorFlow** 기반으로 CNN-LSTM 구축

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement



# Preprocessing

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### 결측치 처리 [1단계]

- <u>팔당댐 데이터(공용량, 방류량)</u>와 <u>강화대교 조위</u> 결측치를 SARIMA로 처리
  - 팔당댐 데이터 : 데이터 변화가 뚜렷하지 않음
  - 강화대교 조위: 주기(약 \*28일)가 명확함

```
for lst in consecutive2:
    for i in col:
    start = lst[0]
    model = pm.auto_arima(df_arima[i][start-n:start], seasonal = True)
    df_arima[i].loc[lst] = model.predict(len(lst)).values
    print(f'{i} 컬럼의 {consecutive2.index(lst)+1}번째 구간 결측치 처리 완료')
    print(df_arima.loc[lst, i].isnull().index)
    print(model.predict(len(lst)))
```

\*출처 : <음력, 양력, 윤달, 윤년>, 전용훈 서울대학교 이학박사

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### 결측치 처리 [2단계]

- <u>청담대교 수위, 잠수교 수위, 한강대교 수위, 행주대교 수위, 행주대교 유량</u> 결측치를 Multivariate CNN-LSTM으로 처리
  - [1단계] 처리 결과(공용량, 방류량)를 Input으로 활용
- · <u>청담대교 유량, 한강대교 유량, 행주대교 수위, 강화대교 조위</u> 결측치를 Multivariate CNN-LSTM으로 처리
  - 잠수교 유량은 결측치 비율이 70.4% → 사용 X

```
def root mean squared error(y true, y pred):
        return K.sqrt(K.mean(K.square(y pred - y true)))
if 'model enc dec cnn' in locals():
    del model enc dec cnn
if 'history' in locals():
    del history
optimizer = tf.keras.optimizers.Adam(lr=0.01) #기본값 0.001
model enc dec cnn = Sequential()
model enc dec cnn.add(Conv1D(filters=64, kernel size=2, padding = 'causal', activation='relu', input shape=(w, xtrain.shape[2])))
model enc dec cnn.add(ConvlD(filters=64, kernel size=2, padding = 'causal', activation='relu'))
model enc dec cnn.add(MaxPooling1D(pool size=2))
model_enc_dec_cnn.add(Flatten())
model enc dec cnn.add(RepeatVector(1))
model_enc_dec_cnn.add(LSTM(200, activation='relu', return_sequences=True))
model_enc_dec_cnn.add(LSTM(200, activation='relu', return_sequences=True))
model enc dec cnn.add(LSTM(200, activation='relu', return sequences=True))
model enc dec cnn.add(LSTM(200, activation='relu', return sequences=True))
model enc dec cnn.add(LSTM(200, activation='relu', return sequences=True))
model_enc_dec_cnn.add(LSTM(200, activation='relu'))
model enc dec cnn.add(Dense(100, activation='relu'))
model enc dec cnn.add(Dense(5))
model enc dec cnn.compile(loss=root mean squared error, optimizer=optimizer, metrics =["accuracy"])
model enc dec cnn.summary()
```

#### 파생변수 생성

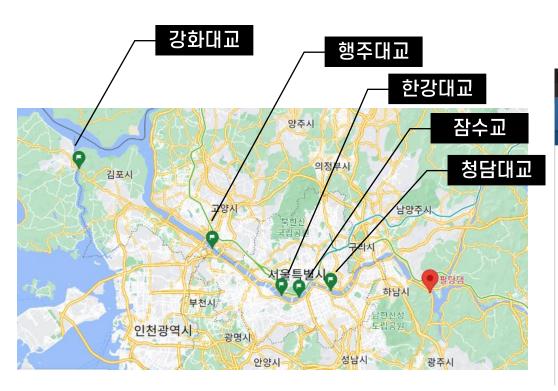
- · <u>강화대교, 팔당댐까지의 거리</u>를 고려하여 새로운 파생변수 생성
  - 230분 전의 강화대교 조위 데이터, 190분 전의 방류량 데이터로 수위 예측
  - 190분과 230분은 상관계수 분석 결과로 선정
  - 연도별 4월 데이터는 <한강홍수통제소> 홈페이지에서 크롤링

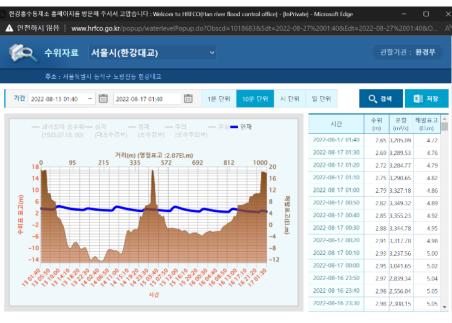
3. Process

2. Data

1. Motivation

- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement





#### 외부 데이터 수집 및 활용

- 서울시 강수량
  - <기상청>에서 제공하는 데이터셋 활용
- 광진교 수위, \*2022년 6월~2022년 7월 한강 주요 다리 수위 데이터
  - <한강홍수통제소> 홈페이지에서 크롤링

2. Data

3. Process

1. Motivation

4. Model

5. Preprocessing

6. Experiments

7. Improvement



활용갤러리 등록

URL 복사

목록 이동

관측소 목록

춘천댐

충주댐

충주조정지댐

팔당댐

평화의댐

한탄강댐

화천댐

횡성댐

한국수력원자력

한국수자원공사

하고수자외공사

한국수력원자력

한국수자원공사

한국수자원공사

한국수력원자력

한국수자원공사

101.400

134.130

64.346

25,010

172.790

47.259

176,400

#### 서울시 강수량 및 강우일수 통계

- 등계개요
- \*통계명:강수량
- \* 통계종류 : 기상청에서 제공하는 서울시 강수량을 제공하는 일반·보고통계 (승인통계 제14102호)
- \* 작성목적 : 서울시 강수량 현황을 제공하여 효율적인 관리 및 재해예방을 위한 기초자료를 제공하는 것을 목적으로 함
- 12 12 10 1C XE 1 12 C
- \* 작성체계: 지역별 기상관서 → 기상청
- \* 공표주기: 정기(매년)
- \*공표범위:지역-서울시

내용 - 서울시 월별 강수량

- 용어설명
- \* 강수량: 비나 눈, 우박 등과 같이 구름으로부터 땅에 떨어져 내린 강수의 양을 말함 어느 기간 동안에 내린 강수가 땅 위를 흘러가거나 스며들지 않고, 땅 표면에 괴어 있다는 가정 아래 그 괸 물의 깊이를 측정한다. 눈 싸락눈 등 강수가 얼음인 경우에는 이것을 녹인 물의 깊이를 측정한다. 비의 경우에는 우량 또는 강우량이라고도 하며, 단위는 빠로 표시

이기타

○출처:기상청

↑ <기상청>에서 제공하는 서울시 강수량 데이터셋



수문 > 실시간 수

↑ <한강홍수통제소> 홈페이지 댐 자료

117.650

1658.074

26.843

211.770

18.663

0.500

674,700

66.906

32,350

873 132

24.167

32.230

2736.616

273.145

343,300

19.885

78.433

60.294

89.477

86.791

0.710

0.185

66,277

76,992

211.670

396.944

37.453

487.380

0.184

1928

0.000

9.621

0.000

0.000

192 146

487.380

3.932

2.768

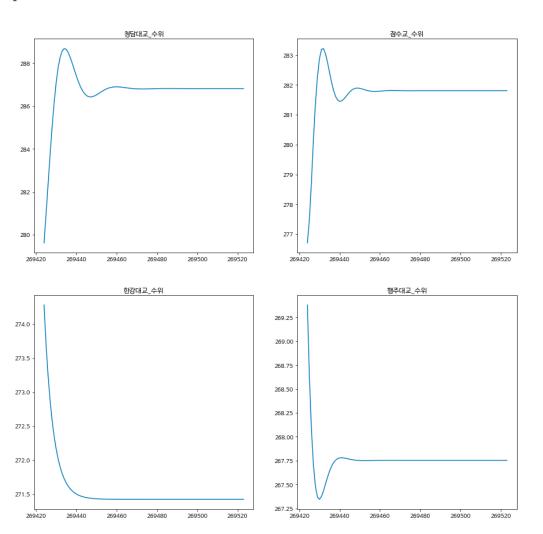
0.000

# Experiments

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### **Baseline Model**

- 전통적인 방법론인 'ARIMA' 모델 활용
- pmdarima.arima의 ARIMA estimator를 활용한 예측 수행



Baseline 코드의 RMSE

청담대교\_수위의 RMSE: 106.28 잠수교\_수위의 RMSE: 90.884 한강대교\_수위의 RMSE: 91.077 행주대교\_수위의 RMSE: 70.401

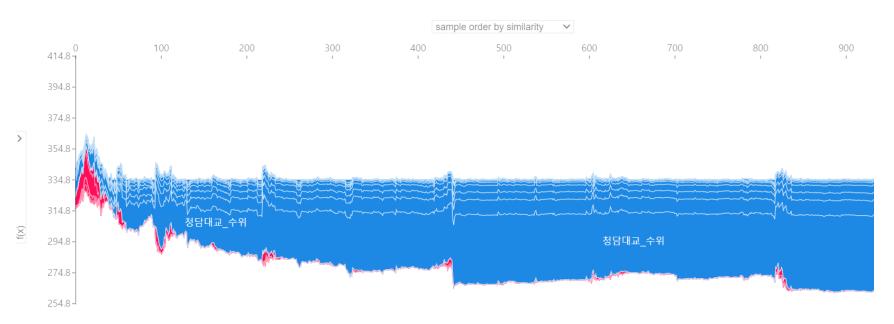
ARIMA를 이용한 한강 다리 수위 예측 결과

#### 1. Motivation

- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### **Machine Learning**

- scikit-learn의 RandomForestRegressor, ExtraTreesRegressor
- xgboost의 XGBRegressor
- Deep Learning 모델보다 성능은 낮았지만, Feature 중요도 파악 가능
  ✓ 이렇게 얻은 인사이트로 Deep Learning 모델 Input 선정에 활용



SHAP을 활용해 청담대교 수위 예측을 위한 XGBRegressor Fitting 결과의 Feature Importance 추출 (1) ※ 파란색 Feature는 Target과 양의 상관관계, 빨간색 Feature는 Target과 음의 상관관계

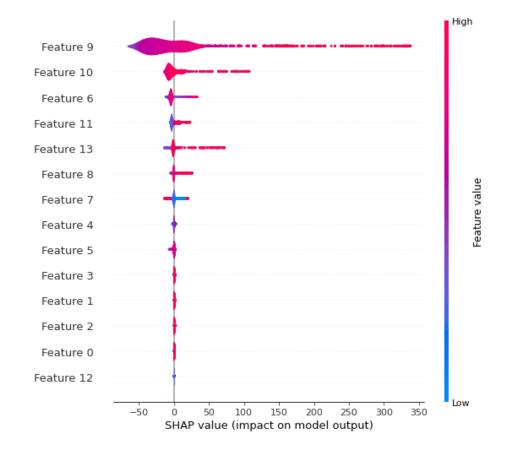
### **Machine Learning**

- scikit-learn의 RandomForestRegressor, ExtraTreesRegressor
- xgboost<sup>Q</sup> XGBRegressor
- Deep Learning 모델보다 성능은 낮았지만, Feature 중요도 파악 가능
   ✓ 이렇게 얻은 인사이트로 Deep Learning 모델 구축에 활용



1. Motivation

- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement



SHAP을 활용해 청담대교 수위 예측을 위한 XGBRegressor Fitting 결과의 Feature Importance 추출 (2)

※ 중요도가 높은 상위 3개의 Feature

- Feature 9 : 청담대교 수위

- Feature 10 : 잠수교 수위

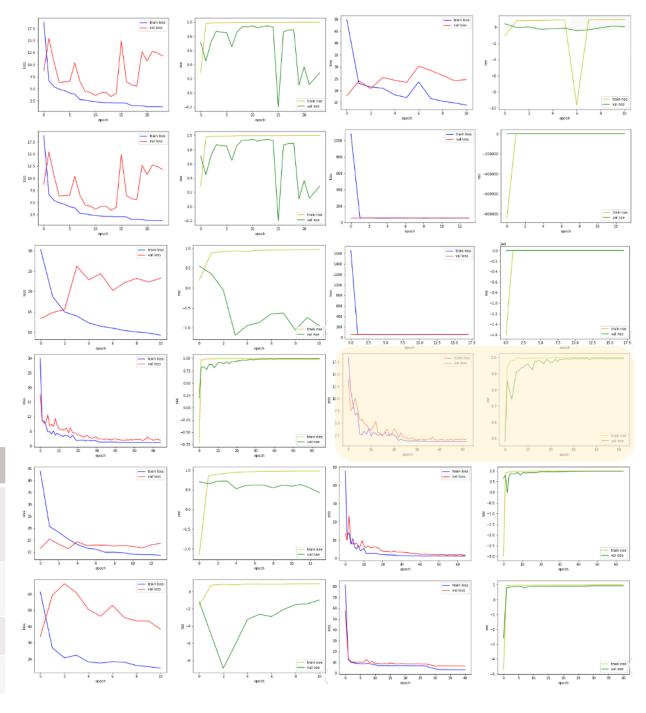
- Feature 6 : 청담대교 유량

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

- 모델 후보
  - 1D CNN
  - LSTM
  - 1D CNN+LSTM (최종 모델)
- 모델별로 하이퍼파라미터 튜닝 & 성능 평가

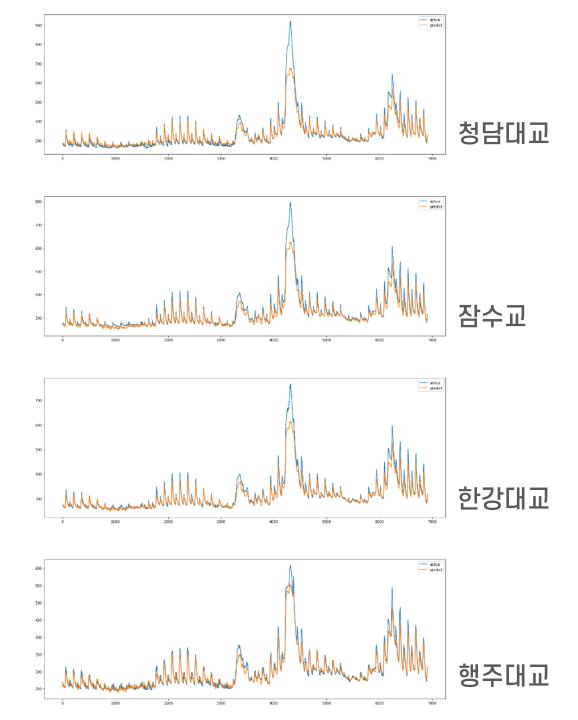
참고) 하이퍼파라미터 튜닝 기준

하이퍼파라미터	튜닝 기준
learning_rate	max 0.01, scheduler 활용 (patience=3)
iteration (batch size)	1000
epoch	100
early_stopping	10



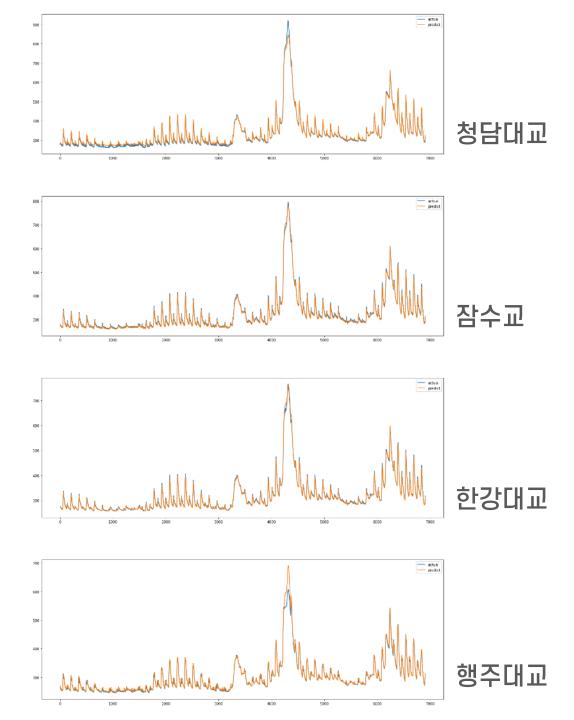
- 1D CNN Visualization
  - Blue: actual
  - Orange: predicted

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement



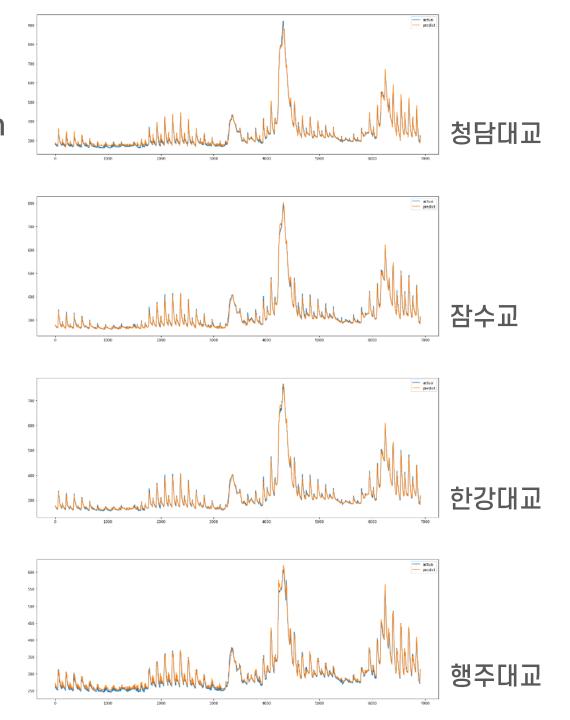
- LSTM Visualization
  - Blue: actual
  - Orange: predicted

- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement



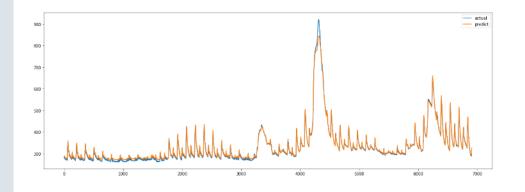
- 1D CNN+LSTM Visualization
  - Blue: actual
  - Orange : predicted

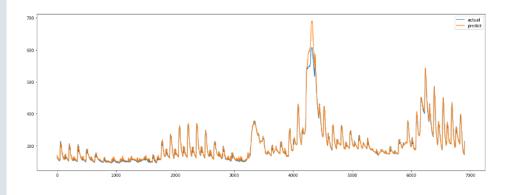
- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement



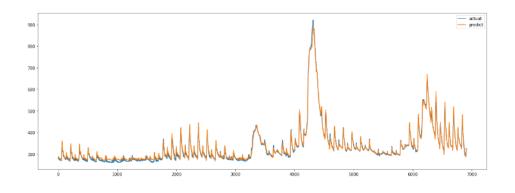
• LSTM vs 1D CNN+LSTM (최종 모델)

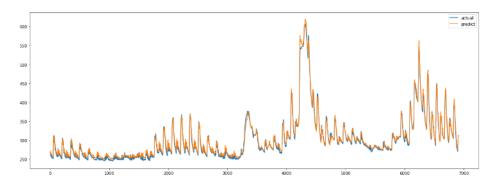
- 1. Motivation
- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement





LSTM 청담대교 예측(위쪽), 행주대교 예측(아래쪽)





최종 모델 청담대교 예측(위쪽), 행주대교 예측(아래쪽)

## Improvement

#### 1. Motivation

- 2. Data
- 3. Process
- 4. Model
- 5. Preprocessing
- 6. Experiments
- 7. Improvement

#### [Limitation]

- •모델의 유용성을 높이려면 Forecast length를 더 길게 잡고 학습해야 함
- 리소스 부족 문제 (데이터셋 활용 문제)

### [Experience]

- 결측치가 적더라도, 전처리 방식에 따라 성능이 크게 달라짐
- •모델 Tuning만큼이나 Dataset 정제도 중요함
- Feature Importance를 확인한 결과, 직관/상식과 맞지 않는 변수 존재했음
- 일반적으로 알려진 이론에 상충하는 결과가 있었음 (ex. BatchNormalization)

#### [Future Work]

- 사전 학습된 모델(pre-trained model) fine-tuning을 통한 성능 향상 기대
- Deep Learning 모델 내부를 살펴볼 수 있는 XAI(설명 가능 AI) 심화 학습
- 한강 수위와 밀접한 관련이 있는 외부 데이터 추가 확보 (지류, 국소지역 강수량 등)

#### 팔당댐 홍수 안전운영에 따른 한강 수위예측 AI 경진대회

알고리즘 | 정형 | 회귀 | 시계열 | RMSE

₩ 상금 : 총 1,100만원

① 2022.07.25 ~ 2022.09.02 17:59 + Google Calendar

**ஃ** 1,161명 📋 D-7

15



팀 팀 멤버 점수 #

.\_\_.YBIGTA.\_\_.



1.29515

## References

- 정성호, 조효섭, 김정엽 and 이기하. (2018). 딥러닝 기반 LSTM 모형을 이용한 감조하천 수위 예측. 한국수자원학회 논문집, 51(12), 1207-1216.
- 조우진, 강동수.(2020).LSTM을 이용한 하천수위 예측기법.한국정보과학회 학술발표논문집,(),439-441.
- 유형주, 이승오, 최서혜, 박문형, Yoo, H., Lee, S. O., Choi, S., & Park, M. (2019). 시계열 자료의 예측을 위한 자료 기반 신경망 모델에 관한 연구: 한강대교 수위예측 적용. 한국방재안전학회논문집, 12(2), 73-82.
- 강지예. "팔당댐 방류량과 서해 조석영향에 따른 한강 하류부 수위상승도달시간 예측." 국내석사학위논 문 한양대학교 대학원, 2010. 서울
- 이정규, 정혜영, 장홍준 (2005) "팔당댐 방류량에 따른 수위상승도달시간 예측", 토목학회 2005 학술발표회논문집.