



Proposal dan Hasil

Implementasi Algoritma ConvLSTM (*Convolutional Long Short Term Memory*) untuk Prediksi Banjir Rob di Kota Semarang

Disusun oleh:

Nama Tim : Nyata dengan Karya

Anggota : Ulil Hidayat¹
: Suwignyo Prasetyo²
: Inlim Ravijai Rumahorbo³

Afiliasi : ¹Stasiun Meteorologi Kalimantan Berau, BMKG
²Pusat Meteorologi Maritim, BMKG
³Stasiun Meteorologi Gunung Silangit, BMKG

PENDAHULUAN

Latar Belakang

Banjir pesisir, yang sering disebut sebagai "banjir rob," secara signifikan mengancam wilayah pesisir, terutama di Kota Semarang (Ismanto et.al, 2022). Model prediksi untuk banjir pesisir dapat membantu dalam sistem peringatan dini dan strategi mitigasi yang efektif, sehingga mengurangi dampak ekonomi dan sosial (Wulandari et.al, 2022). Salah satu metode machine learning yang sering digunakan untuk memprediksi adalah *Convolutional Long Short Term Memory* (ConvLSTM) (Shi et.al, 2015). proyek ini bertujuan untuk memanfaatkan algoritma deep learning ConvLSTM untuk memprediksi kejadian banjir rob di Kota Semarang secara akurat.

Dalam konteks prediksi banjir rob, teknologi kecerdasan buatan (AI) memainkan peran krusial dengan mengintegrasikan berbagai metode untuk meningkatkan akurasi dan efisiensi prediksi. LSTM (*Long Short-Term Memory*) efektif dalam memproses data time-series, seperti data cuaca dan pasang surut, karena kemampuannya dalam menangkap pola temporal yang kompleks dan jangka panjang. Di sisi lain, Convolutional LSTM (ConvLSTM) memperluas kemampuan LSTM dengan menggabungkan struktur konvolusi pada transisi input-ke-state dan state-to-state, yang memungkinkan model ini untuk menangani data spasial secara bersamaan dengan data temporal (Shi et.al, 2015). Dengan memanfaatkan arsitektur ConvLSTM yang menggabungkan struktur konvolusi dan LSTM dalam satu model, dapat dihasilkan prediksi yang lebih akurat untuk fenomena banjir rob dengan mengintegrasikan data geografis dan time-series secara bersamaan. Pendekatan hybrid ini memanfaatkan kekuatan metode AI untuk memberikan prediksi yang lebih komprehensif dan terintegrasi, dibandingkan dengan metode konvensional, sehingga memungkinkan analisis yang lebih cepat dan akurat dalam memitigasi risiko banjir.

Tujuan dan Manfaat

rojek ini bertujuan untuk memanfaatkan algoritma ConvLSTM dalam memprediksi kejadian banjir rob di Kota Semarang. Diharapkan dengan adanya prediksi tersebut, masyarakat dapat memperoleh informasi peringatan dini banjir rob yang akurat, prediksi tersebut dapat menjadi pertimbangan bagi pemerintah daerah dan pihak terkait dalam upaya mitigasi, sehingga dapat mengurangi kerugian ekonomi dan sosial akibat bencana ini.

Cakupan dan Batasan

Project ini menggunakan data selama 5 tahun, yaitu dari 2019 s/d 2023. Metode yang digunakan adalah deep learning dengan pendekatan Convolutional Long Short Term Memory (Conv_LSTM). Parameter yang diambil meliputi curah hujan, suhu, kelembaban relatif, *water level*, arus, angin, tinggi gelombang, dan rekap kejadian banjir rob di Kota Semarang per kecamatan (menyesuaikan grid).

PEMBAHASAN

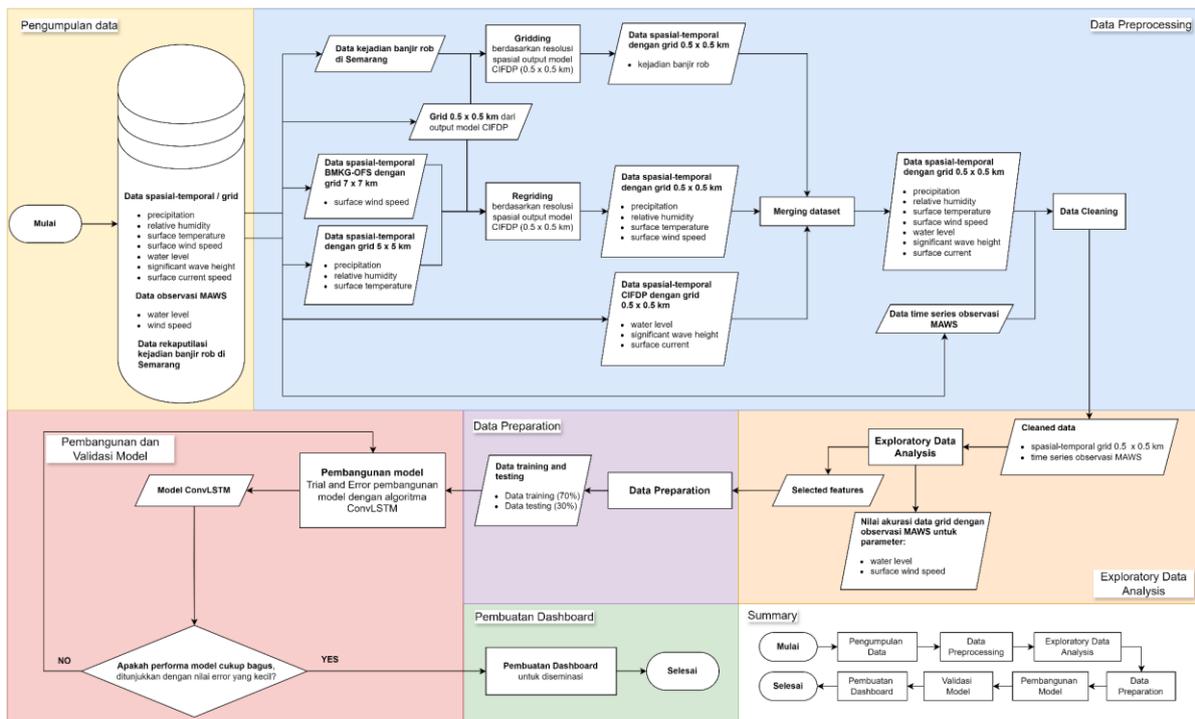
Data

Data yang digunakan dalam project ini terdiri atas data spasial-temporal dan data titik. Detail data yang digunakan dalam pembangunan model prediksi banjir rob di Kota Semarang adalah sebagai berikut.

1. Data kejadian banjir rob setiap kecamatan di Kota Semarang 2019 -2023 yang diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) Kota Semarang dan media online.
2. Data curah hujan, kelembapan dan suhu tahun 2019-2023 dari BMKG yang diperoleh dari panitia dengan resolusi 5 km x 5 km.
3. *Water level, currents, wave height* dan *wind speed* dari Coastal Inundation Forecasting Demonstration Project – Indonesia (CIFDP-I) dengan resolusi 0,5 km x 0,5 km.
4. Data *water level* dan *wind speed* dari pengamatan Marine Automatic Weather Station (MAWS) yang terletak di Pelabuhan Tanjung Mas (lon : -6,94351, lat : 110,43131) tahun 2022-2023. Data tersebut digunakan untuk memverifikasi data model CIFD-I sebelum di input ke dalam model.

Metodologi

Secara umum metodologi dalam pengembangan model ini dilakukan melalui tahapan seperti alur pada Gambar 1.



Gambar 1. Tahapan pengembangan model

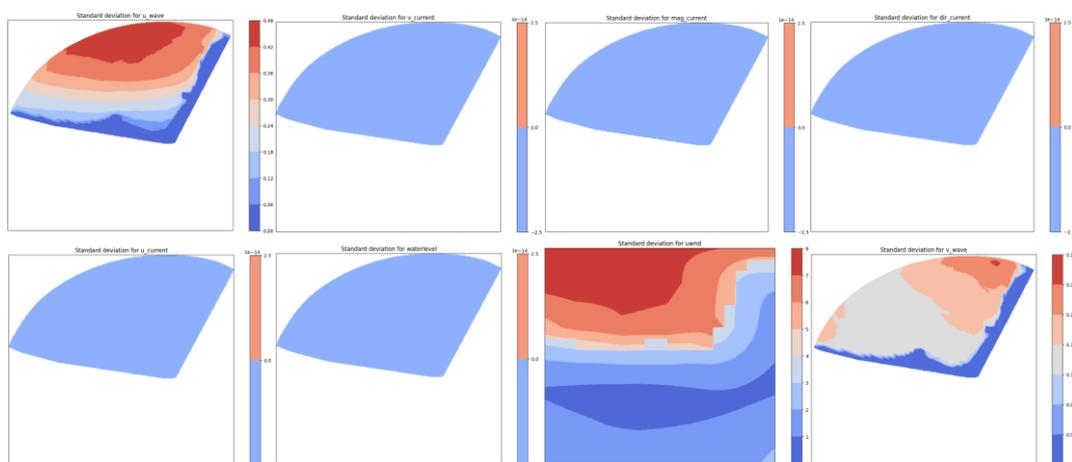
1. Pengumpulan Data

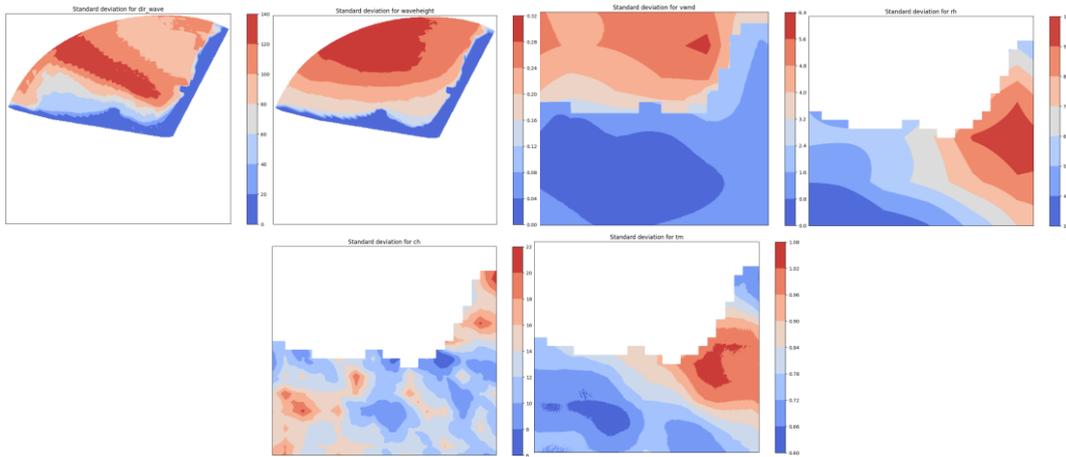
Data kejadian banjir rob per kecamatan di Kota Semarang periode 2019 - 2023 diperoleh dari BPBD Kota Semarang dan media online melalui proses *web scrapping*. Data kejadian tersebut dikumpulkan dalam tabular. Data curah hujan, kelembapan dan suhu periode 2019 -2023 diperoleh dari Panitia. Data *Water level, wind current, wave height* dan *wind speed* model CIFDP-I tahun 2019 -2023 dan data *water level* dan *wind speed* MAWS 2022 -2023 diperoleh dari Pusat Meteorologi Maritim BMKG.

2. Data Preprocessing

- Data kejadian banjir rob per kecamatan dalam bentuk tabular kemudian di konversi ke dalam format grid.
- Data dengan resolusi yang berbeda 5 km x 5 km di di konversi ke resolusi tinggi 0.5 km x 0.5 km dengan metode regrid.
- Data kejadian banjir rob dalam format grid, data curah hujan, kelembapan udara, suhu udara, *water level, current, wave height* dan *wind speed* dengan resolusi 0,5 km di gabungkan dalam satu file netcdf.
- Normalisasi and resampling data untuk memastikan konsistensi data.
- Melakukan perhitungan standar deviasi dan visualisasi data untuk memahami persebaran data.
- Identifikasi korelasi signifikan antar variable prediktor (*curah hujan, kelembapan udara, suhu udara, water level, current, wave height* dan *wind speed*) dan variabel Target(kejadian banjir rob).

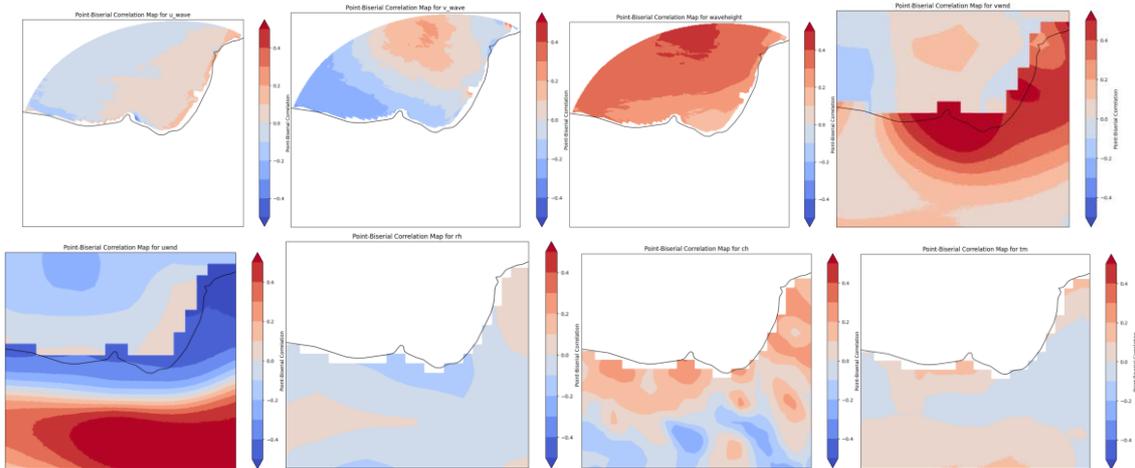
Sebelum melakukan pengembangan model, dilakukan *Explotary Data Analysis* (EDA) untuk mengetahui parameter yang memiliki pengaruh signifikan pada kejadian banjir rob di Kota Semarang, proses ini juga digunakan untuk melihat kualitas data dari data input yang digunakan. Metode yang digunakan dalam proses ini ialah menghitung standar deviasi untuk melihat kualitas data dan menghitung korelasi untuk melihat pengaruh parameter terhadap kejadian banjir rob di Kota Semarang.





Gambar 2. Perhitungan standar deviasi dari setiap parameter

Gambar 2 menunjukkan perhitungan standar deviasi dari setiap parameter, standar deviasi parameter *water level*, *u-current*, *v-current*, *mag current* dan *direction current* menunjukkan nilai standar deviasi nol, artinya parameter tersebut memiliki nilai yang sama yang mengindikasikan kualitas data ini kurang baik. Sedangkan parameter *u wave*, *v wave*, *direction wave*, *waveheight*, *u-wind*, *v-wind*, kelembapan udara, suhu udara dan curah hujan memiliki standar deviasi yang bervariasi, artinya parameter tersebut memiliki nilai yang bervariasi.



Gambar 3. Perhitungan korelasi dari setiap parameter

Gambar 3 menunjukkan perhitungan korelasi dari setiap parameter yang memiliki standar deviasi yang bervariasi. Parameter *u-wave*, *v-wave*, *waveheight*, *u-wind*, *v-wind* dan curah hujan memiliki nilai korelasi yang tinggi, yang artinya memiliki hubungan yang signifikan dengan kejadian banjir rob di Kota Semarang. Sedangkan parameter kelembapan udara dan suhu udara tidak memiliki nilai korelasi yang tinggi, yang artinya tidak memiliki hubungan yang signifikan dengan kejadian banjir rob di Kota Semarang.

Sehingga, diputuskan parameter yang diambil sebagai input model adalah *u wave*, *v wave*, *waveheight*, *u-wind*, *v-wind*, dan curah hujan (6 Parameter).

3. Pengembangan, Hasil, dan Validasi Model

Pengembangan model ini menggunakan metode ConvLSTM dengan menggunakan modul *PyTorch*, dimana proses metode Convolutional secara spasial dan LSTM secara time series secara bersamaan dengan menggunakan module *Conv_LSTM*. Pengembangan model ini menggunakan *training* dan *testing* data input sebanyak 201 hari dan data output (target) sebanyak 67 hari (70% untuk *training* dan 30 % untuk *testing* dari total keseluruhan data yang digunakan). Hal ini dibuat dengan mempertimbangkan *sequence length* yaitu 3 hari. Sehingga model menggunakan inputan data 3 hari belakang (contoh: 3 mei, 4 mei, 5 mei) untuk memprediksi 1 hari kedepan (contoh: 6 mei).

Dikarenakan keterbatasan waktu, dalam tahap ini belum bisa dilakukan *optimize hyperparameters* secara detail (seperti menggunakan teknik *grid search* atau *random search*), sebagai alternatif nya dilakukan *hyperparameter tuning* secara manual berdasarkan nilai akurasi dan *loss per epoch*.

Sebagai awalan digunakan *epoch=10*, *batch_size=10*, *seq_len=3*, *kernel_size=(3,3)*, dan *input_hidden_dim= [16,8,1]*. Berdasarkan performa model, konfigurasi nilai epoch di ubah menjadi 20 (dikarenakan pada *epoch 10* ke 20 masih terdapat pengurangan nilai *loss* walau tidak signifikan) dan *kernel_size* diubah menjadi (5,5) agar bisa mempertimbangkan lebih banyak grid disekitarnya. Konfigurasi model tertera pada Gambar 4.

```
input_dim = input_dim
hidden_dim = [16, 8, 1]
kernel_size = (5, 5)
num_layers = 3
batch_size = 10
seq_len = seq_len
height, width = height, width

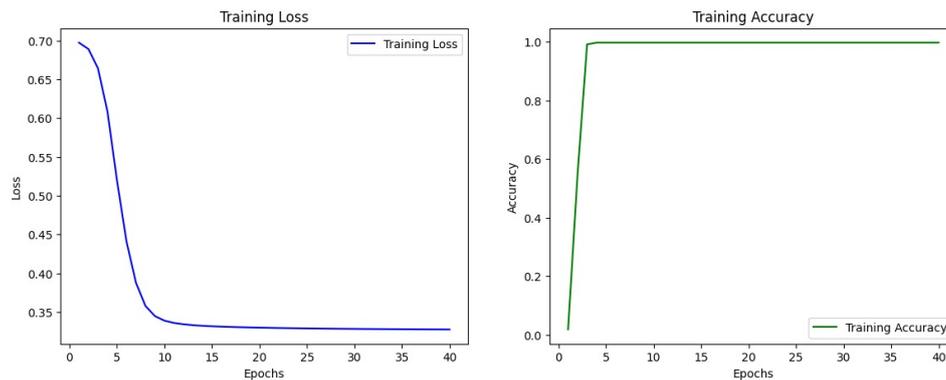
model = ConvLSTM(input_dim=input_dim,
                 hidden_dim=hidden_dim,
                 kernel_size=kernel_size,
                 num_layers=num_layers,
                 batch_first=True,
                 bias=True,
                 return_all_layers=False,
                 )

# Define loss function and optimizer
criterion = nn.BCEWithLogitsLoss() # Binary Cross Entropy with logits
optimizer = optim.Adam(model.parameters(), lr=0.001)
```

Gambar 4. Konfigurasi model

Terdapat masalah ketika dilakukan running model, yakni beberapa parameter yang menjadi data inputan memiliki data NaN. Hal ini membuat model *Conv_LSTM* menjadi error (nilai *loss = nan*). Untuk mengatasi masalah tersebut, dilakukan modifikasi module original *Conv_LSTM* untuk mengabaikan array yang memiliki value NaN dengan metode masking. Detailnya, pertama dibuat *masking_tensor* untuk parameter input, lalu *masking_tensor* tersebut dimasukkan sebagai inputan saat pembangunan model. Hasil akhirnya adalah model dapat berjalan dengan mengabaikan grid NaN.

Untuk performa model ini sendiri masih butuh pengembangan, dimana masih belum berhasil mendeteksi kejadian banjir rob dengan akurat.



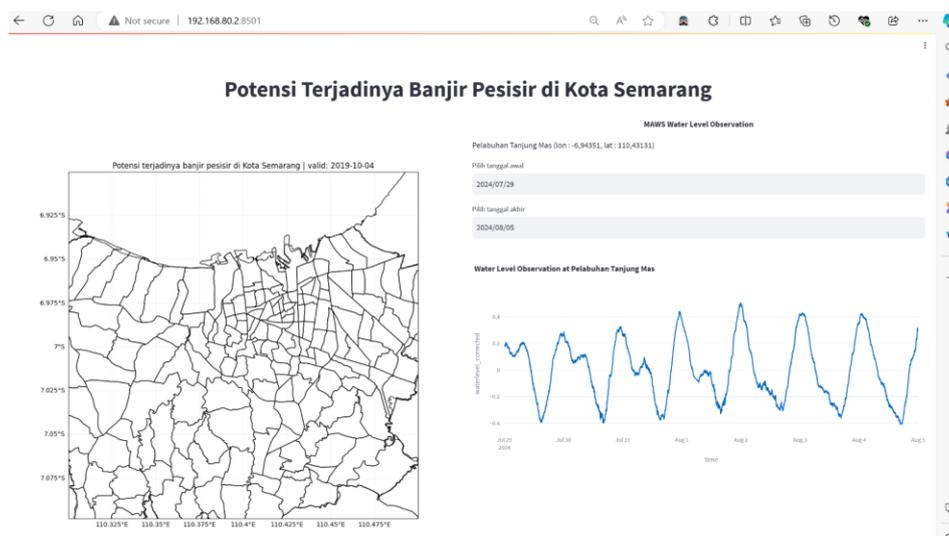
Gambar 5. Grafik epoch terhadap nilai *loss* (kanan) dan terhadap nilai *accuracy* (kiri)

Gambar 5 menunjukkan training *loss* pada *epoch* ke 10-20 stagnan di nilai 0.35. Sedangkan *accuracy* secara umum cukup baik jika hanya melihat nilainya saja, yakni pada *epoch* 0-5 sudah mencapai 90% (0.9). Namun, hal ini disebabkan oleh model yang memprediksi kejadian no rob (tidak ada banjir) untuk semua data test.

Output dari model ini merupakan prediksi probabilitas kejadian banjir rob satu hari kedepan secara spasial dengan resolusi 0,5 km x 0,5 km di Kota Semarang. Output ini ditampilkan dalam area per kelurahan.

4. *Dashbord* Hasil Prediksi Model

Hasil prediksi dari model didiseminasikan dalam bentuk *dashboard* seperti pada Gambar 6. *Dashboard* dibangun menggunakan modul streamlit.



Gambar 6. Tampilan *dashbord* Diseminasi Hasil Prediksi Model

Dashbord diseminasi hasil prediksi model menampilkan peta statis prediksi deterministik kejadian banjir rob satu hari kedepan di Kota Semarang. *Dashbord* juga

menampilkan *timeseries* data pengamatan water level MAWS di Pelabuhan Tanjung Mas selama 7 hari kebelakang.

PENUTUP

Kesimpulan

Projek ini memanfaatkan teknik deep learning ConvLSTM untuk mengembangkan model prediksi banjir rob di Kota Semarang. Model ini menggunakan data parameter *u-wave*, *v-wave*, *waveheight*, *u-wind*, *v-wind* dan curah hujan karena memiliki sebaran nilai yang bervariasi dan hubungan signifikan dengan kejadian banjir rob. Model ini menghasilkan prediksi probabilitas kejadian banjir rob satu hari kedepan secara spasial. Evaluasi model menghasilkan nilai error yang cukup rendah (<0.3) dan akurasi yang cukup tinggi (>0.8). Hasil prediksi dari model dapat didesiminasikan melalui dashboard bersamaan dengan data pengamatan water level MAWS di pelabuhan Tanjung Mas, sehingga hasil prediksi model dapat diakses masyarakat. Diharapkan model ini dapat mengurangi kerugian ekonomi dan sosial yang diakibatkan banjir rob di Kota Semarang.

Potensi pengembangan

Metode ConvLSTM dapat digunakan untuk memprediksi banjir rob secara temporal dan spasial, tidak hanya di Kota Semarang, tetapi juga di wilayah lain di Indonesia yang sering mengalami banjir rob. Metode intinya sama, namun karena data terbatas hanya di Kota Semarang dan Jakarta, data tersebut dapat diperoleh dari sumber lain, seperti Ina-Waves dan Ina-OFS dari Maritim BMKG, serta dari model lain. Model ini juga dapat dimaksimalkan lagi dengan menggunakan data parameter yang berkualitas dan berpengaruh signifikan pada kejadian banjir, seperti parameter pasang surut.

Selain itu, pengembangan juga dapat dilakukan pada proses diseminasi. Penyediaan informasi lebih lengkap seperti prakiraan hujan, pasang surut dan parameter lain akan lebih meningkatkan nilai manfaat dari project ini. Penyebaran informasi kepada masyarakat dan pihak terkait sangat perlu untuk dikembangkan pada project ini. Sarana komunikasi seperti email dapat dikembangkan pada project ini.

Referensi

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28.

https://github.com/ndrplz/ConvLSTM_pytorch