



PREDIKSI CUACA BERDASARKAN KORELASI MULTI- PARAMETER FENOMENA FISIS ALAM DENGAN MACHINE LEARNING SEMINAR PROPOSAL

Ragil Bagus Agung Budiyono - 175090307111003

Pembimbing I

Pembimbing II

(Agus Naba, S.Si., MT., Ph.D)

NIP. 197208061995121001

(Dr.rer.nat. Abdurrouf, S.Si.,M.Si)

NIP. 197209031994121001

Outline

Pendahuluan

Tinjauan Pustaka

Metodologi

Jadwal Kegiatan Penelitian

1

Pendahuluan

● Latar Belakang

○ Cuaca sangat berpengaruh terhadap kehidupan manusia dalam berbagai sektor. Pada pertanian (Cogato er al., 2019), pada pasokan air (Kirono et al., 2016) hingga ke pengendalian pandemi (Tosepu et al., 2020)

Pentingnya cuaca mendorong melakukan peramalan cuaca. Peramalan satu parameter sudah dilakukan seperti peramalan kecepatan angin (Khosravi et al., 2018), prediksi hujan dan temperatur (Kirono et al., 2016)

Penelitian terkait klasifikasi cuaca juga sudah ada, terkait dengan kemampuan kamera untuk mengidentifikasi cuaca secara langsung (Zhang et al., 2016)

● Latar Belakang

○ Berdasarkan latar belakang tersebut, penulis berencana untuk membangun kode program dengan algoritma time-series machine learning dimana prediksi tiap parameter cuaca akan dilakukan dan hasil prediksi tersebut digolongkan menggunakan algoritma classification machine learning, sehingga didapatkan ramalan cuaca multi parameter. Dalam penelitian kedepannya juga akan diteliti terkait hubungan antar parameter fisis yang paling mempengaruhi sehingga tercipta kategori cuaca pada waktu tersebut

- Tujuan

- 1. Mengetahui korelasi antara parameter fenomena fisis terhadap cuaca yang terjadi.
 2. Mengetahui cara untuk memprediksi cuaca terhadap multi-parameter fenomena fisis dengan metode regresi dan klasifikasi.
 3. Mengetahui kemampuan model terhadap korelasi tiap parameter dan perubahan hasil apabila parameter ditambahkan atau dikurangi.
 4. Mengetahui kemampuan model terhadap prediksi tahunan pada rentang waktu dan parameter terlibat yang ditentukan.

● Batasan Masalah

- 1. Cuaca yang diprediksi memiliki kriteria cerah, Sebagian mendung, mendung, hujan, cerah disertai hujan, sebagian mendung disertai hujan
2. Cuaca yang diprediksi berdasarkan data curah hujan tiap jam di kota Malang pada Januari 2010 – Januari 2021
3. Parameter faktor penentu hujan secara fisis yang digunakan adalah suhu minimum-maksimum, kecepatan dan arah angin, presipitasi dan kelembapan.

2

Tinjauan Pustaka

● Cuaca

- Cuaca bermakna keadaan atmosfer pada satuan waktu tertentu
- Cuaca memiliki parameter yang diungkapkan, seperti suhu, tekanan, angin, kelembapan dan lainnya
- Karena cuaca sangat mempengaruhi kehidupan manusia, banyak metode peramalan dibangun

(Wirjohamidjojo & Swarinoto, 2010; Watts, 2014)

● Parameter Cuaca

- Parameter secara umum adalah temperature minimum, temperature maksimum, keadaan angin, keadaan awan dan evaporasi.

(Potter & Coleman, 2003)

● Parameter Cuaca

- Variabel utama yang diambil untuk fenomena fisis adalah volume udara atau air di atmosfer, massa jenis, temperatur, dan tekanan. Dari variabel tersebut umumnya para peneliti cuaca (meteorologis) menggunakan beberapa model matematis untuk menjelaskan fenomena fisis tersebut, diantaranya persamaan konservasi momentum Navier-Stokes, hukum pertama termodinamika terkait konservasi energi, persamaan konservasi massa udara, persamaan kontinuitas massa uap air dan persamaan gas ideal. Persamaan tersebut dapat menjadi acuan korelasi antar parameter, tetapi korelasi antar parameter pastinya beragam bergantung pada faktor lain, dan hal tersebut menyebabkan prediksi menggunakan persamaan matematis memiliki hasil yang tidak optimal

(Lions et al., 1992; Potter & Coleman, 2003)

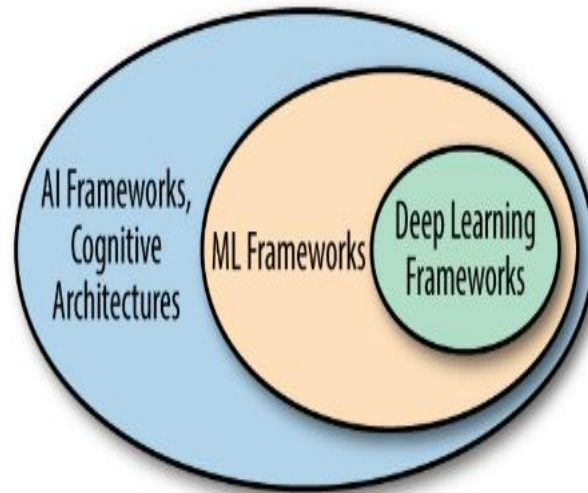
● Artificial Intelligence dan Machine Learning

○ Cakupan AI sangat luas, mencakup penyelesaian regresi, klasifikasi, *clustering* hingga ke penalaran, perencanaan dan navigasi. Ide besar dari AI adalah mampu menyelesaikan kebutuhan manusia yang mendasar dengan baik atau bahkan *lebih* baik

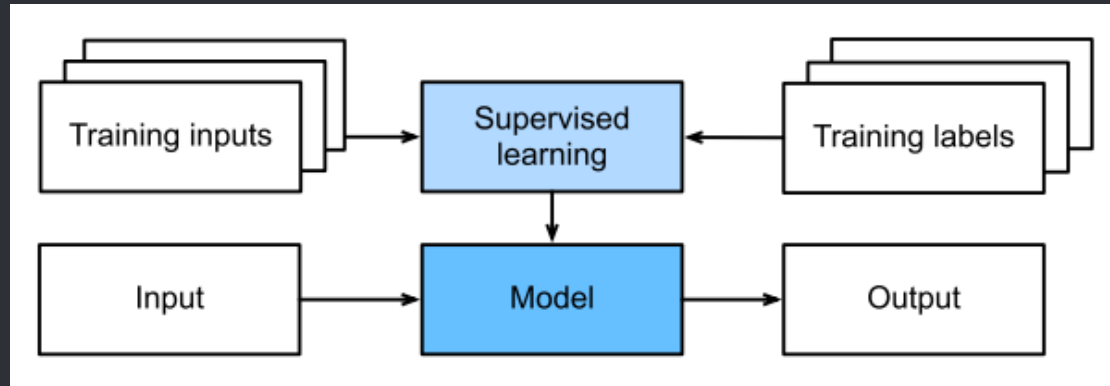
- *Machine Learning* (ML) merupakan algoritma naungan AI. Fungsionalitas awalnya adalah mengubah data mentah menjadi sesuatu yang bernilai, hal tersebut berlaku pada ML secara konvensional. Dengan penelitian yang matang, ML pada terobosan terbarunya mampu benar-benar mengekstrak sebuah pola pada data mentah dalam berbagai jenis data.

(Morgan, 2018)

- Artificial Intelligence dan Machine Learning



- Supervised Learning



(Czum, 2020)

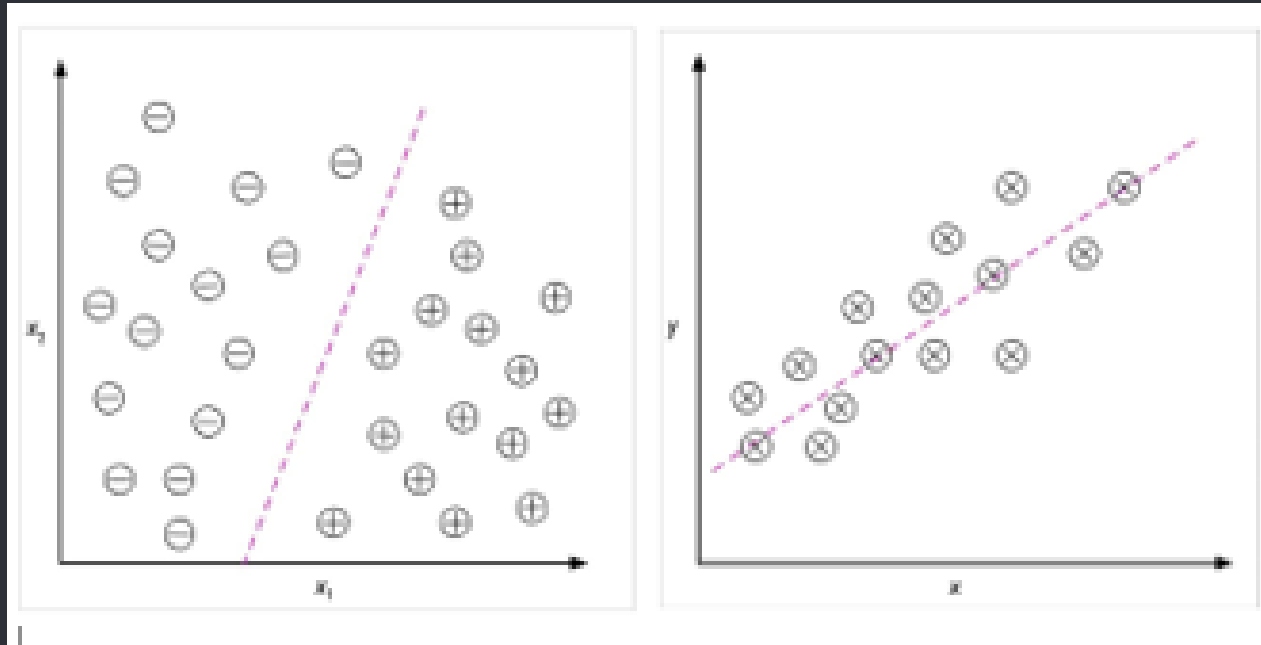
● Regresi dan Klasifikasi

Sederhananya regresi merupakan prediksi yang dilakukan dengan angka real berdasarkan data yang ada. Regresi dapat menambahkan data yang kosong maupun meneruskan data yang sudah ada. Contoh regresi adalah prediksi harga rumah tahunan, prediksi harga saham, prediksi curah hujan. Kunci dari regresi adalah keluaran yang berupa nilai kontinu, sehingga sangat cocok untuk menyelesaikan permasalahan prediksi.

- Lain halnya klasifikasi, ML akan mengelompokan data sesuai dengan kategori label yang telah ditentukan (pada kasus supervised learning). Keluaran dari klasifikasi adalah nilai diskrit yang menentukan label dari data. Contohnya adalah bagaimana klasifikasi kucing dan anjing, kanker ganas dan tidak, serta klasifikasi keadaan cuaca.

(Raschka & Mirjalili, 2017; Czum, 2020).

- Regresi dan Klasifikasi

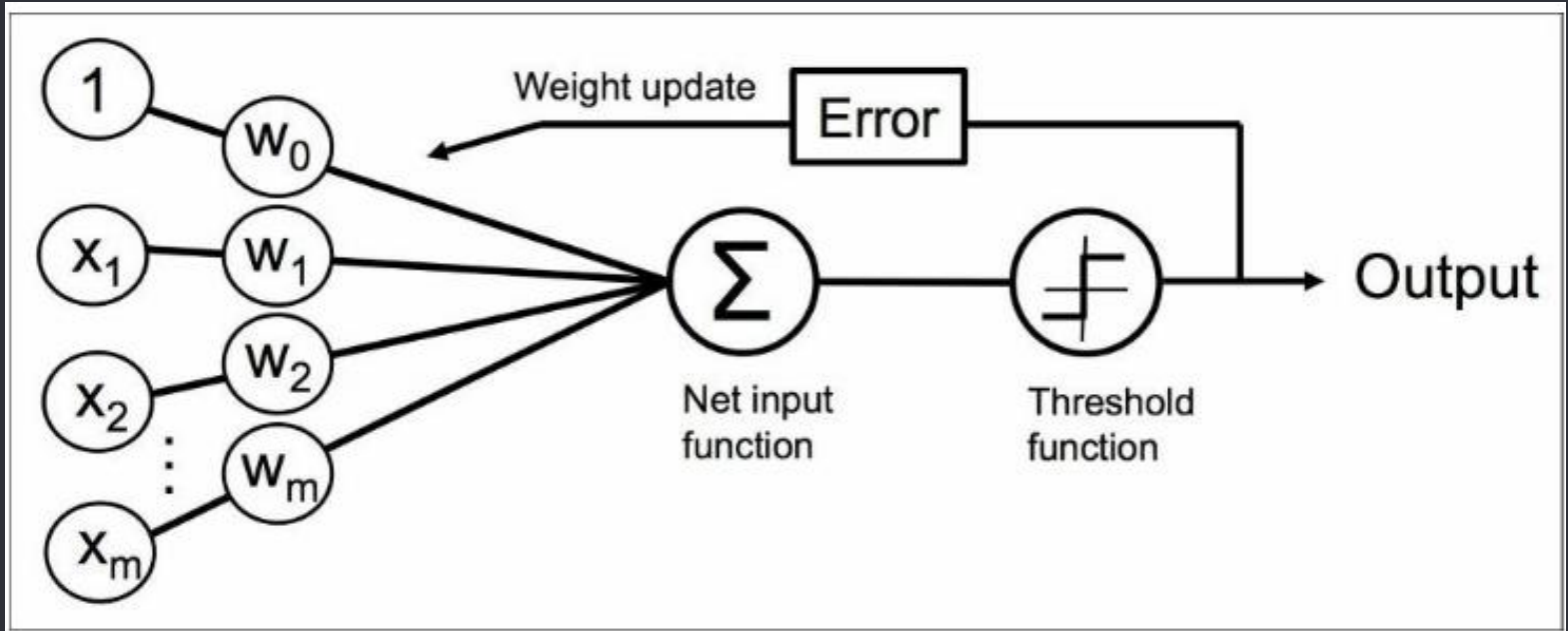


- Neural Network

○ NN dibangun berdasarkan penggambaran sistem syaraf manusia, ketika beragam *input* masuk lalu diproses untuk menghasilkan keluaran biner. NN akan mempelajari data yang diberikan lalu memberikan *update* bobot sehingga model akan semakin baik, maka dari itu NN merupakan algoritma yang dilakukan secara berulang (*epoch*) untuk menentukan bobot yang sesuai dengan label keluaran. Persamaan update bobot:

$$w_j := w_j + \Delta w_j$$
$$\Delta w_j = \eta (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}$$

- Neural Network



● Time Series Neural Network

- NN punya kecenderungan untuk menemukan pola pada sebuah data masukan. Penemuan pola tersebut didasarkan pada *input* sebelumnya, sehingga data yang telah di proses akan mempengaruhi bobot ke depannya. Permasalahannya adalah NN konvensional cenderung “lupa” apabila data yang dipanggil terlalu lama, dan NN konvensional tidak sensitif terhadap waktu. Karena itu, NN konvensional kurang baik untuk mengelola data dengan waktu sensitif seperti ramalan curah hujan dan harga saham.

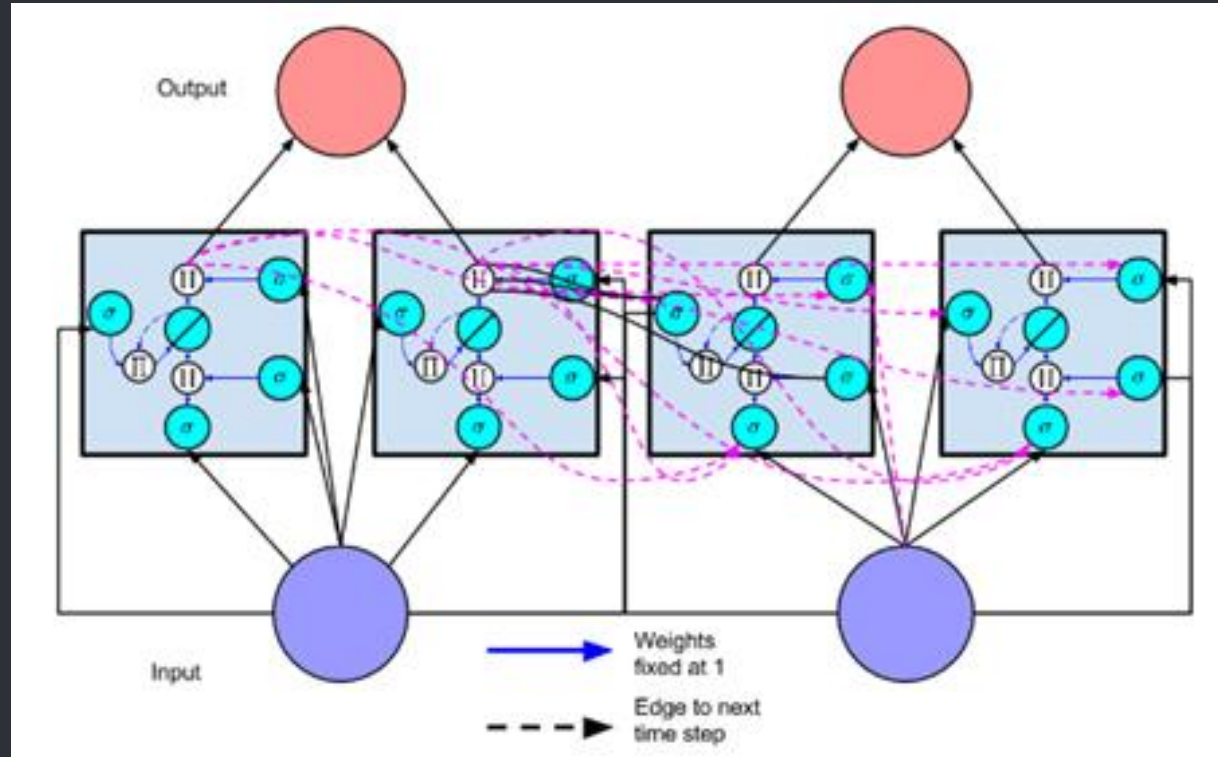
(Lipton et al., 2015)

● Time Series Neural Network

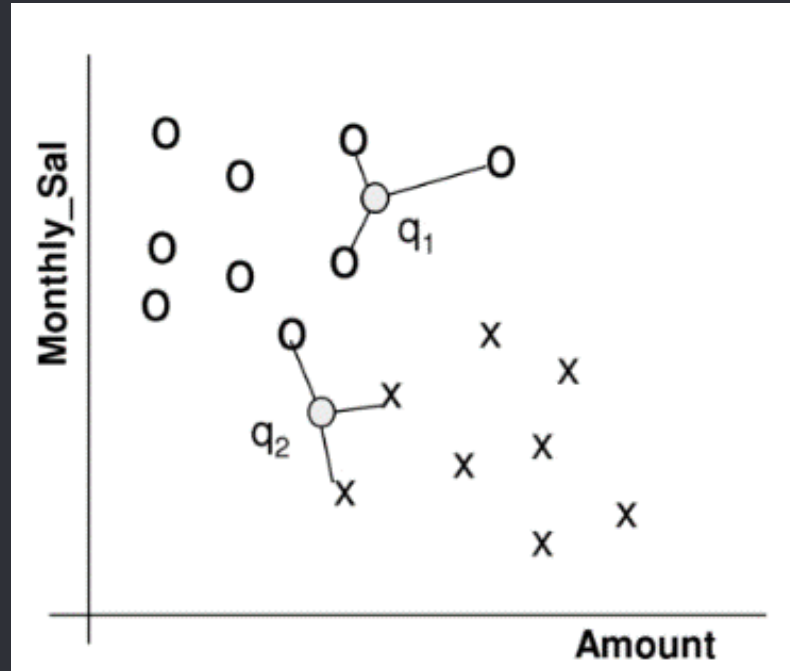
- Recurrent NN sebenarnya adalah modifikasi dari feed-forward Neural Network dengan menambahkan komponen waktu. Recurrent NN mampu mengingat data yang telah lampau dengan urutan yang baik. Kelemahannya adalah Recurrent NN tidak dapat memilih konstanta mana yang harus diingat dan dilupakan, sehingga muncul metode baru yaitu Long-Short Term Memory (LSTM).

(Lipton et al., 2015)

Time Series Neural Network



- K-Nearest Neighbors



● K-Nearest Neighbors

- Poin utama dari algoritma ini adalah indentifikasi sebuah poin di data dan melihat data tetangga sekitarnya. Poin tersebut akan diidentifikasi berdasarkan jarak antara tetangga terdekatnya. Persamaannya adalah sebagai berikut:

$$d(q, x_i) = \sum_j w_j \delta(q_j, x_{if})$$

● Loss Function

○ *Loss function* merupakan fungsi untuk menentukan seberapa besar kesalahan yang didapat dari *training* maupun *fitting* ML. Pada regresi, umumnya menggunakan *root mean square error* (RMSE) atau menggunakan *mean absolute error* (MAE).

- RMSE dan MAE digunakan berdasarkan kebutuhan, dan jenis data yang digunakan. Sebagian merujuk bahwa MAE memiliki keunggulan untuk menunjukkan *loss function* yang stabil sehingga dapat merepresentasikan data dengan benar, tetapi tidak sedikit pula yang masih menggunakan RMSE sebagai standar *loss function*

(Chai & Draxler, 2014)

● Python

- Tergolong bahasa tingkat tinggi, Python merupakan salah satu bahasa yang efektif digunakan untuk melakukan penelitian numerik. Walaupun tidak se-efisien Fortran atau C++, Python memiliki kelebihan lain dengan banyaknya *library* yang mendukung banyak penelitian, seperti penggunaan NumPy, Pandas, TensorFlow, dan lain sebagainya.

(Dubois et al., 2007).

3

Metodologi

- **Waktu dan Tempat Pelaksanaan**

- Penelitian ini dilakukan pada bulan Maret 2021 sampai bulan Mei 2021 di Jalan Srigading Dalam No. 58K Kota Malang.

● Alat dan Bahan

1. Sebuah Laptop Asus dengan Spesifikasi Processor Intel i7 770HQ CPU @ 2.8 GHz, 4 core(s), 8 Logical Processor(s), RAM DDR4 16.00 GB, SSD 512 GB, NVIDIA Geforce GTX 1050.
2. Sebuah *Operating System (OS)* Windows 10 Pro
3. Beberapa *library* Python yaitu NumPy, Pandas, Matplotlib, TensorFlow, Keras, Seaborn dan Scikit-Learn.
4. NVIDIA-CUDA untuk membuka akses penggunaan GPU pada *library* TensorFlow.
5. IDE (*Integrated Development Environment*) berupa Spyder dengan versi instalasi Python 3.8.3.
6. Data cuaca tiap jam yang didapatkan dari laman penyedia data cuaca dengan alamat: <https://www.visualcrossing.com/weather/weather-data-services#/>

● Tahapan Penelitian

1. Persiapan Komputasi
2. Pengolahan Data
3. Pembuatan *Model Neural Network* dan *K-Nearest Neighbors*
4. Pelatihan *Neural Network* dan *K-Nearest Neighbors*
5. Pengujian *Neural Network* dan *K-Nearest Neighbors*
6. Evaluasi dan Visualisasi ML

● Pengolahan Data

1. Data dibersihkan dengan menghilangkan parameter yang tidak memiliki nilai
2. Merubah data menjadi time-series data dengan mengatur time-step. Misalkan time-step diatur selama 12 jam, maka akan ada 12 data uji kemudian data ke 13 menjadi label keluaran untuk dibandingkan hasilnya, lalu data bergeser dari 2-13 dengan data ke 14 menjadi label keluaran dan seterusnya
3. Data diskalakan dengan range 0-1

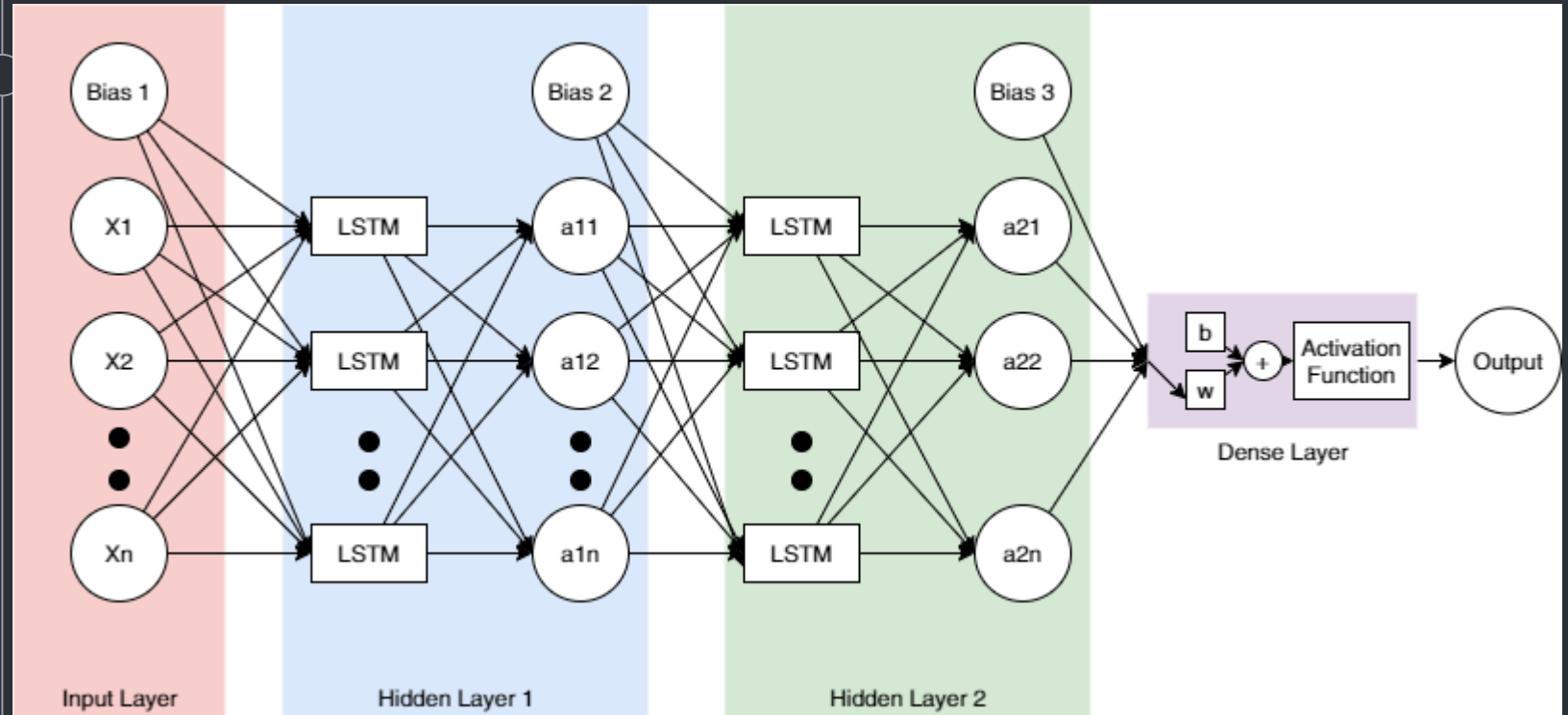
- **Pengolahan Data**

No	Input					Target				
	t-12	t-11	t-10	t-9	t+8	t+9	t+10	t+11
1	250	277	119	143	11.7	54	62	89
2	277	119	143	151	54	62	89	509
3	119	143	151	212	62	89	509	153
...
...
323	165	446	246	668	38.7	0.1	45.9	131.8
324	446	246	668	190	0.1	45.9	131.8	489.2
325	246	668	190	87.1	45.9	131.8	489.2	276.3

● Pembuatan Model

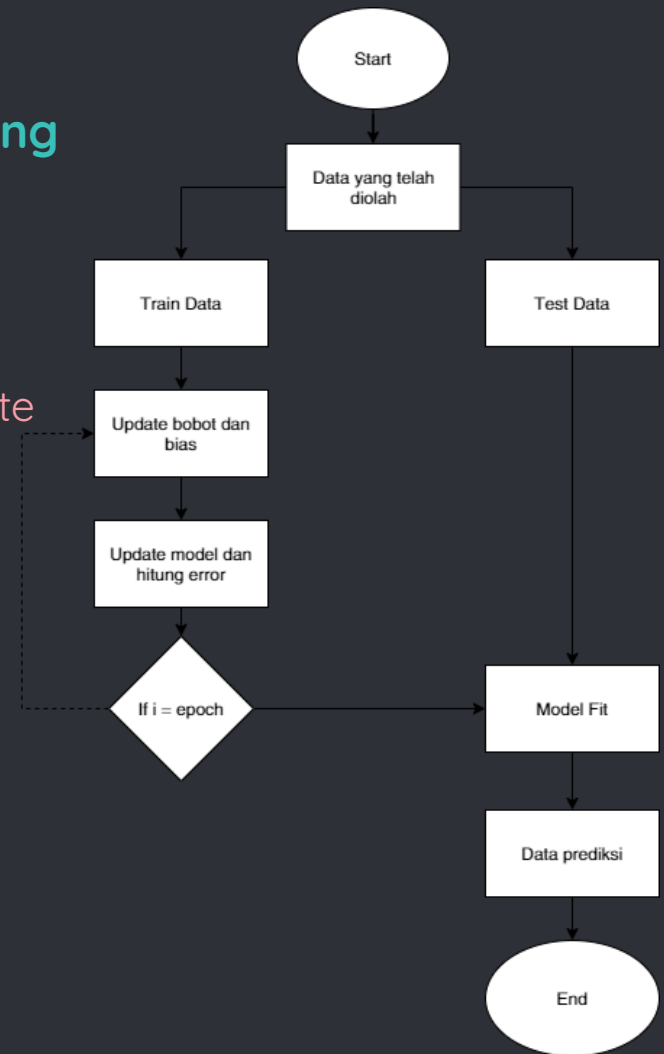
1. Untuk regresi, unit dan layer dibuat dengan bentuk seperti time-series regresi
2. *Hyperparameter* seperti banyaknya unit dan layer, *loss function*, *optimizer*, *learning rate* dan *activation function* akan disesuaikan tiap *training* dilaksanakan. Nilai awal yang di masukan sebelum penyesuaian adalah 1 *layer input*, 2 *hidden layer* LSTM, dan 1 *output layer*. Unit pada *hidden layer* pertama sebanyak 64, dan pada *hidden layer* kedua sebanyak 32. *optimizer* yang digunakan adalah 'adam' dengan opsi *learning rate* 'default'. *Loss function* yang digunakan adalah RMSE dan MAE, kemudian dicari mana yang terbaik. *Activation function* yang digunakan adalah sigmoid dan relu, kemudian dicari mana yang terbaik
3. Model pada K-Nearest Neighbors dapat merujuk pada persamaan sebelumnya. K-Nearest Neighbors tidak memiliki arsitektur khusus seperti Neural Network karena tidak memiliki neuron

● Pembuatan Model



Pelatihan dan Pengujian Machine Learning

1. Data di split menjadi dua
2. Train dijalankan hingga epoch selesai
3. Selama dijalankan bobot dan bias di update
4. Setelah model jadi maka diuji di data test



● Evaluasi dan Visualisasi Machine Learning

- Evaluasi dilakukan dengan mengamati loss function dan hasil prediksi, apabila kurang memuaskan maka hyperparameter dapat diubah-ubah
- Pengujian lain dengan mencoba kombinasi parameter cuaca hingga didapatkan kombinasi terbaik, lalu dicari tahu alasan fisisnya
- Visualisasi digunakan untuk melihat perbandingan data tes dan train serta korelasi antar parameter

4

Jadwal Kegiatan Penelitian

● Jadwal Kegiatan Penelitian

No	Kegiatan	Waktu									
		2021									
		3			4			5			
1	Pembuatan Proposal										
2	Persiapan Komputasi										
3	Pengolahan Data										
4	Pembuatan Model ML										
5	Pelatihan Model ML										
6	Pengujian Model ML										
7	Evaluasi dan Visualisasi ML										
8	Penulisan Tugas Akhir										

“

Terimakasih