

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Jung min han, Yu qian ang, Ali malkawi, Holly W. samelson pada tahun 2021 melakukan penelitian dengan judul “*Using recurrent neural networks for localized weather prediction combined use of public airport data and on-site measurement*” untuk memprediksi parameter cuaca: suhu, kelembaban, kecepatan dan arah angin ,dataset yang digunakan diperoleh dari National Solar radiation Database (NSRDB) dan dapat di akses di www.nsrdb.nrel.gov serta data hasil pengamatan cuaca selama satu tahun. Data dari bulan Januari hingga juni akan digunakan sebagai data train sedangkan data Juli hingga Desember akan digunakan sebagai data prediksi, model yang digunakan untuk memprediksi adalah LSTM dan GRU dan evaluasi menggunakan MSE. Hasil dari penelitian menunjukkan bahwa GRU memiliki performa terbaik daripada LSTM untuk melakukan prediksi suhu, kecepatan angin, arah angin, kecuali kelembaban [11] .

Penelitian yang dilakukan oleh shiva nosouhian, Fereshteh nosouhian dan Abbas kazemi khoshouei pada tahun 2021 dengan judul “*A Review of recurrent neural network architecture for sequence learning : Comparison between LSTM and GRU*”. Mereka melakukan penelitian menggunakan LSTM dan GRU dengan menggunakan dataset motion sequences sebanyak 1307, dataset dibagi menjadi *train*, *validation* dan *test* dengan komposisi 0.7, 0.2 dan 0.1, kemudian dilakukan

standarisasi data dengan melakukan *rescale* dataset sehingga mempunyai *input feature* dengan rentang yang sama. Hasil dari penelitian ini di evaluasi menggunakan *loss* dengan menggunakan konfigurasi 20 dan 45 epochs menunjukkan bahwa GRU mempunyai performa yang lebih baik [12].

Rui fu, zuo zhang dan li li pada tahun 2016 memprediksi arus lalu lintas menggunakan model LSTM, GRU dan ARIMA, dataset yang digunakan didapat dari Caltrans Performance Management System (PeMS), serta hasil dari prediksi di evaluasi menggunakan *mean square error* (MSE) dan *mean absolute error* (MAE). Dari hasil penelitian memperlihatkan bahwa LSTM dan GRU memiliki performa yang lebih unggul dari ARIMA, dan GRU memiliki kinerja yang lebih unggul dari LSTM [9].

Penelitian selanjutnya dilakukan oleh Peter T yamak, Li yujian dan Pius K gadosey pada tahun 2019 dengan judul “*A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting*” menggunakan dataset harga bitcoin yang didapatkan dari www.cryptodatadownload.com tanggal 28 November 2014 sampai 5 Juni 2019. Preprocessing data dilakukan menggunakan MinMax Normalization dan mengubah data menjadi stasioner menggunakan Augmented Dicky Fuller (ADF). Hasil dari penelitian didapatkan bahwa ARIMA memiliki nilai yang terbaik dari proses evaluasi MAPE dan RMSE, selanjutnya diikuti oleh GRU dan kemudian LSTM [10].

Eko Supriyadi pada tahun 2019 melakukan penelitian yang berjudul “Prediksi Parameter Cuaca Menggunakan *Deep Learning Long Short Term Memory* (LSTM)”

dengan menggunakan parameter cuaca suhu udara, kelembaban udara, kecepatan angin, dan tekanan udara selama bulan Januari dan Februari 2019 yang berasal dari pengamatan sinoptik Stasiun Meteorologi Maritim Tanjung Priok, pada penelitian ini diterapkan algoritma *deep learning* LSTM dengan jumlah layer 200 menggunakan perbandingan training serta data test 9 : 1, dan kemudian dilakukan pengukuran nilai RMSE dan RMSE *update*, Diperoleh nilai RMSE *update* untuk parameter suhu, kelembaban, kecepatan angin, dan tekanan udara masing-masing bernilai 0.576, 2.8687, 2.1963, dan 1.0647. Dari parameter tersebut hanya parameter suhu dan kelembaban udara yang mengalami pertambahan RMSE seiring bertambahnya waktu [13].

Rafidah atika, Arna fariza dan Ali ridho barakbah melakukan penelitian dengan judul “*Forecast Rainfall data time series using attribute long short-term memory*” yang meneliti mengenai prakiraan curah hujan menggunakan LSTM dan *Single Moving Average*, data yang digunakan merupakan data curah hujan BMKG selama 11 (sebelas) tahun dari tiga alat penakar hujan yang terdapat di Surabaya yang bersumber dari www.dataonline.bmkg.go.id. hasil dari nilai prediksi selama 1 tahun dibandingkan dengan data aktual menggunakan nilai RMSE dan *Mean Square Error* (MSE), dari hasil penelitian menunjukkan bahwa metode LSTM mampu memprediksi dengan lebih baik dibandingkan dengan *Single Moving Average* [3].

Xiao dkk melakukan penelitian pada tahun 2019 yang berjudul “*Short and Midterm Sea surface temperature prediction using time-series satellite data and LSTM-*

AdaBoost combination approach”, para peneliti menggunakan suhu permukaan laut yang berasal dari *National Oceanic and Atmospheric Administration* (NOAA), selanjutnya dilakukan prediksi menggunakan arsitektur model LSTM, AdaBoost, Support Vector Regression (SVR), *Back Propagation Neural Network* (BPNN) dan S-LSTM-AdaBoost yang kemudian hasil prediksi dibandingkan dengan metode statistik RMSE, MAE, Pearson’s dan Kernel Density estimation (KDE). Pada penelitian ini diketahui bahwa semua model menghasilkan nilai yang baik untuk prediksi 1-10 hari kedepan[14].

Penelitian selanjutnya yang dilakukan oleh Muhammad rizki, Setio Basuki dan Yufis azhar pada tahun 2020 yang berjudul “Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang”. Data yang digunakan diperoleh dari www.dataonline.bmkg.go.id dari tahun 1989 – 2017, atribut yang digunakan yaitu curah hujan kota malang yang kemudian dilakukan pengujian prediksi menggunakan arsitektur LSTM dengan konfigurasi optimal hidden layer 256, jumlah epoch optimal yaitu 150 dan komposisi optimal data train 50% dan test 50% memiliki tingkat error yang paling rendah [4].

Penelitian yang dikembangkan oleh Ali khumaidi, Ridwanraafiudin dan Indra permana solihin pada tahun 2020 yaitu mengenai “Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung” dari data *time series* hasil pengumpulan data manual oleh peneliti yaitu parameter PM10, ISPU, suhu dan kelembaban. Pemodelan LSTM dengan 4 *Hidden Layer*, menentukan jumlah

ukuran batch (yaitu, 32), menentukan pengoptimal untuk epoch Adam bernilai 1000, dan menentukan fungsi *loss* dengan MSE. Hasil pemodelan LSTM menunjukkan bahwa jaringan berkinerja cukup baik saat melatih dan menguji prediksi. Pemodelan memberikan akurasi prediksi yang cukup baik untuk tiga parameter (suhu, kelembaban dan ISPU). Hal ini ditunjukkan dengan nilai RMSE yang diprediksi lebih kecil dari nilai standar deviasi kumpulan data uji. Namun hasil prediksi yang dihasilkan dari keempat parameter pengujian didasarkan pada urutan terbaik yaitu kelembaban, suhu, ISPU, dan PM10 [15].

Hanifa dkk melakukan penelitian mengenai “Perbandingan Metode LSTM dan GRU (RNN) Untuk Klasifikasi Berita Palsu Berbahasa Indonesia” pada tahun 2021, penelitian tersebut mengklasifikasi berita bohong berbahasa Indonesia dengan menggunakan model RNN yaitu LSTM dan GRU, data yang digunakan yaitu berasal dari portal berita online Indonesia seperti: Detik, Tribun, Liputan 6 dll sebanyak 600 berita dengan total berita valid 600 dan berita palsu 228. Nilai akurasi yang dihasilkan oleh model sebesar 73 % menggunakan LSTM dan 64 % menggunakan GRU [6].

Penelitian yang dilakukan oleh Yaseen essa, Hugh G.P hunt dan Ritesh ajoodha pada tahun 2021 yang berjudul “*Short-Term Prediction Of Lightning In Southern Africa Using Autoregressive Machine Learning Techniques*”, pada penelitian menggunakan data petir Cloud To Ground (CG) selama tahun 2018 yang diperoleh dari The South African Weather Service, pembagian data sebesar 70% / 30% untuk *train* dan *test*. Tujuan penelitian ini memprakirakan kejadian petir selam periode per tiga jam

menggunakan tiga metode yaitu LSTM, ARIMA dan AR dan kemudian di evaluasi menggunakan *Mean Absolute percentage error* (MAPE) dan *Root Mean Square Error* (RMSE), dari hasil penelitian didapatkan bahwa ketiga metode memiliki nilai *error* yang tinggi dan disarankan oleh peneliti untuk menggunakan model *non-parametrics* kedepannya [16].

Ramadhan dkk dalam penelitian yang berjudul “*Forecasting of Sea Level Time Series using RNN and LSTM Case Study in Sunda Strait*” yang bertujuan untuk memprediksi tinggi muka air laut menggunakan data tinggi muka air laut per jam selama 3 (bulan) yang berasal dari IDSL (Inexpensive Device Fore Sea Level Measurement), metode yang digunakan adalah membandingkan hasil prediksi RNN dan LSTM selama 3, 5, 7, 10 dan 14 hari. Dari hasil prediksi LSTM memberikan hasil yang lebih baik dibandingkan RNN dengan nilai koefisien korelasi R^2 0.97 dan nilai RMSE 0.036 untuk prakiraan selama 14 hari [17].

Penelitian yang dilakukan oleh Tita latiffia, Putu wira buana dan Ni kadek dwi rusjyanthi pada tahun 2022 dengan judul “Model prediksi cuaca menggunakan metode LSTM” menggunakan data curah hujan dan suhu periode tahun 2013-2020, pada penelitian ini dilakukan konfigurasi *epoch* sebanyak 50, 100, 200, 400 dan *batch size* yaitu 32, 50. Model prediksi cuaca menggunakan metode LSTM dengan akurasi terbaik rata-rata diperoleh dengan menggunakan *batch size* 100 serta epoch 50 sedangkan nilai RMSE dan MAPE terbaik diperoleh yaitu 1.7444 dan 1.9499% [5].

Penelitian selanjutnya yang dilakukan Mochammad agus sholeh dan Rahmat hidayat pada tahun 2022, tentang “Perbandingan Model LSTM dan GRU Untuk Memprediksi Harga Minyak Goreng Di Indonesia”. Dataset untuk prediksi adalah data harga minyak selama 3 (tiga) tahun terakhir yang berasal dari Pusat harga Pangan Strategis Indonesia (PIHPS), data yang terkumpul diproses terlebih dahulu dengan mengubah tipe data dan mengisi data harian yang kosong dengan nilai dari hari sebelumnya, data yang telah diproses sebelumnya dimodelkan dengan LSTM dan GRU, pada penelitian ini model LSTM memiliki performa yang lebih unggul dibandingkan model GRU dengan MAE *validation* nilai ± 0.04 dan MSE *validation loss* bernilai ± 0.002 [7].

Yadi karyadi dan Handri santoso (2022) dalam penelitiannya yang berjudul “Prediksi Kualitas Udara Dengan Metoda LSTM, *Bidirectional* LSTM, dan GRU”, menggunakan dataset kualitas udara kota Bandung pada tahun 2019 dengan parameter suhu, kelembaban, Particular matter PM10 dan Indeks Standar Pencemar Udara (ISPU). Dari hasil penerapan algoritma dilakukan evaluasi menggunakan RMSE dari data *testing* dan dibandingkan dengan standar deviasi, LSTM dan LSTM Bidirectional menunjukkan hasil yang lebih bagus dibandingkan model GRU [8]

Tabel 2. 1 Penelitian Terkait

Peneliti, Tahun	Judul Penelitian	Jumlah dan Nama Atribut	Arsitektur Model	Preprocessing	Metode Evaluasi	Lokasi Dataset	Sumber Dataset	Nilai Evaluasi
Jung min han, Yu qian ang, Ali malkawi, Holly W. samelson (2021) [11]	<i>Using recurrent neural networks for localized weather prediction combined use of public airport data and on-site measurement</i>	1. Suhu 2. Kelembaban 3. Kec. Angin 4. Arah angin	<i>Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU)</i>	-	<i>Mean Square Error (MSE)</i>	Bandar udara internasional Logan, Boston	www.nsrdb.nrel.gov/archives	MSE LSTM Suhu : 8.72 MSE LSTM Kelembaban : 170.7 MSE LSTM kec.angin : 1.34 MSE LSTM arah angin : 3515 MSE GRU Suhu : 3.38 MSE GRU Kelembaban : 181.42 MSE GRU kec.angin : 0.99 MSE GRU arah angin : 3025
Shiva nosouhian, Fereshteh nosouhian, Abbas kazemi khoshouei (2021) [12]	<i>A review of recurrent neural network architecture for sequence learning : Comparison between LSTM and GRU</i>	1. Motion sequence	<i>Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU)</i>	<i>Rescale, Standarization</i>	Loss	-	-	Multi LSTM loss : 0.0012 - 0.0411 Single LSTM loss : 0.006 – 0.0343 Multi GRU loss : 0.007 – 0.014 Single GRU loss : 0.008 – 0.0187
Rui fu, zuo zhang dan li li (2016) [9]	<i>Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction</i>	1. Jumlah arus lalu lintas per 30 detik	<i>Long Short Term Memory (LSTM), Gated</i>	-	<i>Mean square error (MSE) dan Mean absolute error (MAE)</i>	Oakland, California	-	ARIMA MSE : 841.0065 ARIMA MAE : 19.11753 LSTM MSE : 710.0502 LSTM MAE : 18.127758

			<i>Recurrent Unit (GRU) dan Autoregressive Moving Average (ARIMA)</i>					GRU MSE : 668.930 GRU MAE : 17.2116
Peter T Yamak, Li Yujian, Pius K Gadosey (2019) [10]	<i>A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting</i>	1. Harga bitcoin	<i>Autoregressive Moving Average (ARIMA), Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU)</i>	<i>Augmented dickey fuller (ADF), MinMax Normalization</i>	<i>Mean absolute percentage error (MAPE) dan Root mean square error (RMSE)</i>	-	www.cryptodata.com/download.com/data/	ARIMA MAPE : 2.76 ARIMA RMSE : 3012.53 LSTM MAPE : 6.80 LSTM RMSE : 603.68 GRU MAPE : 3.97 GRU RMSE : 381.34
Eko supriyadi (2019) [13]	Prediksi Paramater cuaca menggunakan <i>Deep Learning Long-short term Memory (LSTM)</i>	1. Suhu udara, 2. Kelembaban udara, 3. Kecepatan angin 4. Tekanan udara	<i>Long Short Term Memory (LSTM)</i>	Standarisasi data menggunakan <i>Mean</i> dan <i>Standar Deviasi</i>	<i>Root mean square error (RMSE)</i>	Tanjung priuk, Jakarta Utara	-	<i>RMSE update</i> RMSE suhu : 0.576 RMSE kelembaban : 2.8687 RMSE kec. angin : 2.1963 RMSE tekanan udara : 1.0647
Rafidah Atika, Arna Fariza, Ali Ridho	<i>Forecast Rainfall data time series using attribute long short-term memory</i>	1. Curah Hujan	<i>Long Short Term Memory (LSTM), single</i>	<i>Data Cleaning</i>	<i>Root mean square error (RMSE) dan Mean square error (MSE)</i>	Surabaya, Jawa Timur	www.dataonline.bmkg.go.id/akses_data	RMSE LSTM : 13.766 – 15.719 MSE LSTM : 189.514 – 247.090 RMSE <i>single Moving Average</i> : 13.545 – 24.360

Barakbah (2019) [3]			<i>Moving Average</i>					RMSE <i>single Moving Average</i> : 183.455 – 593.418
Changjiang Xiao, Nengcheng Chen, Chuli hu, Ke Wang, Jianya Gong, Zeqiang Chen (2019) [14]	<i>Short and Mid-term sea surface temperature prediction using satellite data and LSTM-AdaBost combination approach</i>	1. <i>Sea Surface Temperature</i>	<i>Long Short Term Memory (LSTM), AdaBoost, Support vector regression (SVR), Backpropagation neural network model(BPNN) S-LSTM-AdaBoost</i>	<i>Deseasonalization dan Normalization</i>	<i>Root mean square error (RMSE) dan Mean absolute error (MAE)</i>	Laut Cina timur	www.ncdc.noaa.gov/oiss	RMSE LSTM : 0.01 – 1.99 RMSE Adaboost : 0.92 – 3.58 MAE LSTM : -0.01 – 3.73 MAE Adaboost : 1.05 – 4.93
Muhammad Rizki, Setio Basuki, Yufis Azhar (2020) [4]	Implementasi Deep Learning Menggunakan Arsitektur <i>Long Short Term Memory</i> Untuk Prediksi Curah Hujan Kota Malang	1. Curah hujan	<i>Long Short Term Memory (LSTM)</i>	MinMaxScaler	<i>Root mean square error (RMSE)</i>	Malang, Jawa Timur	www.dataonline.bmkg.go.id/akses_data	RMSE <i>data train</i> : 12.079 RMSE <i>data test</i> : 11.288
Ali Khumaidi, Ridwan Raafi'udin,	Pengujian Algoritma <i>Long Short Term Memory</i> untuk prediksi kualitas	1. Suhu 2. Particulate matter (PM10) 3. Kelembaban	<i>Long Short Term Memory (LSTM)</i>	Normalisasi dengan cara konversi tanggal dan ke	<i>Root mean square error (RMSE) dan</i>	Bandung, Jawa Barat	-	RMSE suhu :3.15 RMSE PM10 : 1.76 RMSE Kelembaban : 6.97 RMSE ISPU : 1.85

Indra Permana Solihin (2020) [15]	udara dan suhu kota Bandung	4. Indeks standar pencemar udara (ISPU)		format <i>Datetime</i>	Standar Deviasi			Standar Deviasi suhu : 4.40 Standar Deviasi PM10 : 1.75 Standar Deviasi Kelembaban : 9.32, Standar Deviasi ISPU : 1.90
Aini Hanifa, Sugih Ahmad Fauzan, Muhammad Hikal, Muhammad Bahrul Ashfiya (2021)[6]	Perbandingan Metode LSTM dan GRU (RNN) untuk Klasifikasi Berita Palsu Berbahasa Indonesia	1. Berita	<i>Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU)</i>	<i>Stemming, Tokenizing, Penghapusan Stopword, converting ke number sequence, padding</i>	Akurasi	-	Portal berita online : www.liputan6.com, www.detik.com, www.tribunnews.com, www.turnbackhoax.id	Akurasi LSTM : 73 % Akurasi GRU : 64 %
Yaseen Essa, Hugh G.P. Hunt, Ritesh Ajooda (2021) [16]	<i>Short Term Prediction of lightning in south Africa using Autoregressive Machine Learning</i>	1. Data petir Cloud To Ground (CG)	<i>Long Short Term Memory (LSTM), Autoregressive Moving Average (ARIMA), Auto Regressive (AR)</i>	<i>Dickey fuller test</i>	<i>Mean absolute percentage error (MAPE) dan Root mean square error (RMSE)</i>	Afrika Selatan	-	RMSE AR : 8579 RMSE ARIMA : 8301 RMSE LSTM : 9426 MAPE AR : 15312 MAPE ARIMA : 15080, MAPE LSTM : 3705

Annas Wahyu Ramadhan, Didit Aditya, Deni Saepudin, Indwiarti, Semeidi Husrin, Adiwijaya (2021) [17]	<i>Forecasting of sea Level time series using RNN and LSTM case study in sunda Strait</i>	1. Tinggi muka air laut	<i>Recurrent Neural Network (RNN) dan Long Short Term Memory (LSTM)</i>	<i>Transforming data</i>	<i>Root mean square error (RMSE) dan Coefficient correlation (CC)</i>	Pulau Sebesi, Lampung	-	RMSE LSTM : 0.036 CC LSTM : 0.97 RMSE RNN : 0.047 CC RNN : 0.96
Tita Lattifia, Putu Wira Buana, Ni kadek Dwi Rusjyanthi (2022) [5]	Model Prediksi Cuaca Menggunakan Metode LSTM	1. Suhu 2. curah hujan	<i>Long Short Term Memory (LSTM)</i>	Denormalisasi	<i>Root mean square error (RMSE) dan Mean absolute percentage error (MAPE)</i>	-	-	RMSE : 1.744 MAPE : 1.9499 %
Mochammad Agus Sholeh, Rahmat Hidayat (2022) [7]	Perbandingan Model LSTM dan GRU Untuk Memprediksi Harga Minyak Goreng Di Indonesia	1. Harga minyak goreng	<i>Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU)</i>	Transformasi data, Transposisi	<i>Mean absolute error (MAE) dan Mean square error (MSE)</i>	Seluruh Provinsi Indonesia	www.hargapan.id/tabel-harga/pasar-tradisional/daerah	LSTM MAE : ± 0.04 LSTM MSE : ± 0.002
Yadi Karyadi, Handri	Prediksi Kualitas Udara Dengan Metoda LSTM,	1. Suhu 2. Kelembaban 3. Particular Matter (PM10)	<i>Long Short Term Memory (LSTM),</i>	Konversi data ke format <i>Datetime</i>	<i>Root mean square error (RMSE)</i>	Bandung, Jawa Barat	https://docs.google.com/uc?export=download&id=1cgfee	RMSE LSTM : 1.84 – 6.96 RMSE LSTM-bi: 1.86 – 7.07 RMSE GRU : 1.76 – 6.96

Santoso (2022) [8]	<i>Bidirectional LSTM</i> dan GRU	4. Indeks standar pencemar udara (ISPU)	<i>Bidirectional LSTM</i> dan <i>Gated Recurrent Unit (GRU)</i>				qeIkINiTp5XSfPgCF2_3SOUsVJ8	
M. Devid Alam Carnegie (2022)	Prediksi Curah Hujan Menggunakan <i>Long Short Term Memory</i> dan <i>Gated Recurrent Unit</i>	<ol style="list-style-type: none"> 1. Penyinaran matahari 2. Kelembaban (jam 07, 13, 18) 3. Tekanan 4. Suhu (max, min) 5. Kecepatan angin 6. curah hujan 7. keadaan cuaca 	<i>Long Short Term Memory (LSTM)</i> , dan <i>Gated Recurrent Unit (GRU)</i>	<i>Data cleaning, Data Scaling, Data transformation, Data Split,</i> Bahasa Pemrograman menggunakan Python	<i>Root mean square error (RMSE), Mean square error (MSE), Mean absolut deviation (MAD), Confusion Matrix</i>	Kotabumi, Lampung	https://drive.google.com/drive/folders/1W0JjGF4AG2eGk9ey4XNS0g_BGAWHw7YK	<p>RMSE LSTM 1: 16.81 RMSE GRU 1: 19.23 RMSE LSTM 2: 19.50 RMSE GRU 2: 19.41 RMSE LSTM 3: 21.10 RMSE GRU 3: 22.20</p> <p>MSE LSTM 1: 282.55 MSE GRU 1: 369.82 MSE LSTM 2: 380.18 MSE GRU 2: 376.66 MSE LSTM 3: 445.34 MSE GRU 3: 492.89</p> <p>MAD LSTM 1: 10.43 MAD GRU 1: 11.05 MAD LSTM 2: 11.00 MAD GRU 2: 11.16 MAD LSTM 3: 10.65 MAD GRU 3: 12.03</p> <p>Akurasi LSTM 1: 49% Akurasi GRU 1: 62% Akurasi LSTM 2: 54% Akurasi GRU 2: 61% Akurasi LSTM 3: 70%</p>

								<p>Akurasi GRU 3: 67%</p> <p>Precision LSTM 1: 46% Precision GRU 1: 58% Precision LSTM 2: 55% Precision GRU 2: 59% Precision LSTM 3: 78% Precision GRU 3: 64%</p> <p>Recall LSTM 1: 51% Recall GRU 1: 66% Recall LSTM 2: 46% Recall GRU 2: 73% Recall LSTM 3: 55% Recall GRU 3: 75%</p> <p>F1 score LSTM 1: 48% F1 score GRU 1: 62% F1 score LSTM 2: 50% F1 score GRU 2: 65% F1 score LSTM 3: 65% F1 score GRU 3: 69</p>
--	--	--	--	--	--	--	--	--

2.2 Perbedaan penelitian yang diajukan dengan penelitian sebelumnya

Perbedaan penelitian ini dan penelitian sebelumnya terdapat pada arsitektur model yang digunakan, pada penelitian sebelumnya prediksi curah hujan hanya dilakukan menggunakan algoritma LSTM [3] [4] [5], maka dari itu pada penelitian ini ditambahkan satu (1) arsitektur model lagi yaitu GRU. Pada penelitian ini running LSTM dan GRU akan menggunakan beberapa input, istilah ini kita sebut dengan “multivariate”, perbedaan dengan pendekatan "univariate" di mana hanya satu atribut yang digunakan sebagai input, pendekatan "multivariate" memperhitungkan beberapa variabel sekaligus. Dengan menggunakan data multivariate, model LSTM dan GRU dapat mempelajari korelasi dan interaksi antara atribut tersebut. selanjutnya kedua model tersebut dibandingkan nilai evaluasinya untuk mengetahui model manakah yang terbaik dalam melakukan prediksi curah hujan di Stasiun Geofisika Lampung Utara.

2.3 Landasan Teori

2.3.1 Curah Hujan

Hujan adalah sebuah bentuk presipitasi yang berbentuk cair maupun padat, sedangkan curah hujan ialah air hujan yang jatuh ke permukaan bumi dengan asumsi permukaan tersebut tidak menguap, meresap dan mengalir. Satuan curah hujan dinyatakan dalam millimeter (mm) yang diukur menggunakan penakar hujan observatorium dan penakar hujan tipe Hellman, Nilai yang menyatakan sifat hujan adalah:

1. Banyaknya air hujan (R), atau banyaknya salju secara kumulatif selama waktu sampai pada saat pengamatan; umumnya digunakan satuan millimeter. Curah hujan (R) 1 mm menyatakan banyaknya air hujan yang tertampung pada bidang permukaan datar pada luasan 1 m² sebanyak 1 liter;
2. Kelebatan/intensitas hujan (I), yakni banyaknya hujan atau salju tiap satuan waktu selama hujan atau selama salju turun (r); nilainya dinyatakan dalam mm/detik, $I = r/t$;
3. Hari hujan (H), yakni banyaknya hari dengan ada hujan dalam jumlah tertentu (biasanya digunakan ketetapan dalam sehari ada hujan sebanyak 0,5 mm atau lebih). Nilainya dinyatakan dalam H/minggu, H/bulan, dan per waktu lain yang digunakan [18].

Intensitas curah hujan merupakan jumlah curah hujan yang terjadi pada periode waktu tertentu, BMKG sendiri membagi kriteria intensitas curah hujan menjadi 5 kriteria.

Tabel 2. 2 Kriteria Intensitas Hujan

NO	Kriteria Hujan	Intensitas Curah Hujan Per Jam	Intensitas Curah Hujan Per Hari
1	Sangat Ringan	< 1 mm	< 5 mm
2	Ringan	1 - 5 mm	5 – 20 mm
3	Normal	5 – 10 mm	20 – 50 mm
4	Lebat	10 – 20 mm	50 – 100 mm
5	Sangat Lebat	> 20 mm	> 100 mm

Tabel 2. 3 Kriteria Intensitas Hujan Per Bulan

NO	Kriteria Hujan	Intensitas Curah Hujan Per Bulan
1	Rendah	0 - 100 mm
2	Menengah	101 – 300 mm
3	Tinggi	301 – 500 mm
4	Sangat Tinggi	> 500mm

Sifat hujan merupakan perbandingan jumlah curah hujan yang terjadi selama satu bulan di suatu tempat dengan nilai rata-rata atau normalnya pada bulan dan tempat yang sama, sifat hujan dibagi menjadi 3 kriteria yaitu:

1. Atas Normal (AN) : Jika Nilai perbandingannya > 115%
2. Normal (N) : Jika Nilai perbandingannya antara 85% - 115%
3. Bawah Normal (BN) : Jika nilai perbandingannya < 85%

2.3.2 Sifat Umum Hujan

Hujan berkaitan dengan jenis awan. Hujan dengan intensitas tinggi tetapi waktu hujannya pendek umumnya berasal dari awan jenis Cumulus atau Cumulonimbus. Hujan dengan intensitas rendah umumnya berasal dari awan jenis Stratus. Hujan yang berlangsung lama umumnya berasal dari awan Altostratus tebal atau dari Nimbostratus [18].

Di kawasan lintang tengah dan lintang tinggi:

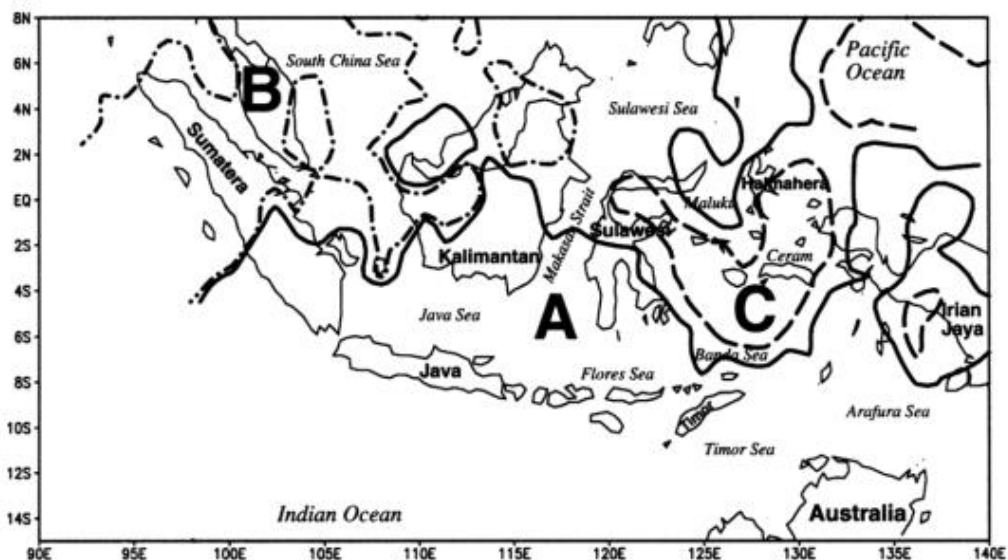
1. Hujan sering dalam bentuk salju atau es;
2. Hujan umumnya terjadi di daerah perenggan dan siklon;
3. Intensitasnya rendah tetapi sering berlangsung lama.

Di kawasan tropis:

1. Umumnya hujan dalam bentuk air, sangat jarang berbentuk es;
2. Intensitasnya besar, tetapi waktu hujannya pendek;
3. Di atas daratan umumnya hujan terjadi pada siang atau sore hari;
4. Di daerah pegunungan umumnya hujan terjadi pada sore dan menjelang malam hari
5. Di daerah pantai hujan sering terjadi di malam atau menjelang pagi hari.

2.3.3 Pola Curah Hujan

Terdapat 3 (tiga) pola curah hujan di Indonesia, dimana pembagian pola tersebut di bagi berdasarkan wilayah di Indonesia [19]

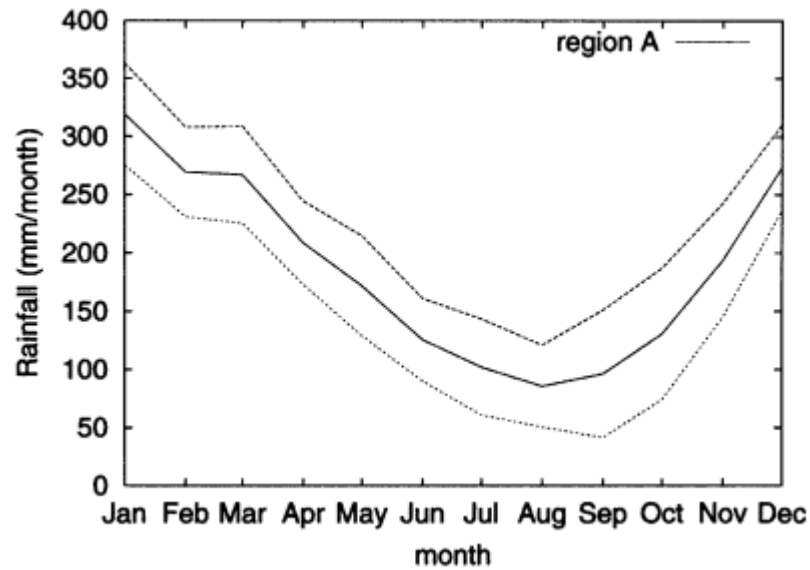


Gambar 2. 1 Pola Curah Hujan di Indonesia

1. Wilayah A

Pola Curah Hujan wilayah A memiliki satu puncak curah hujan yang terjadi pada bulan november hingga bulan maret dan curah hujan terendah terjadi pada bulan

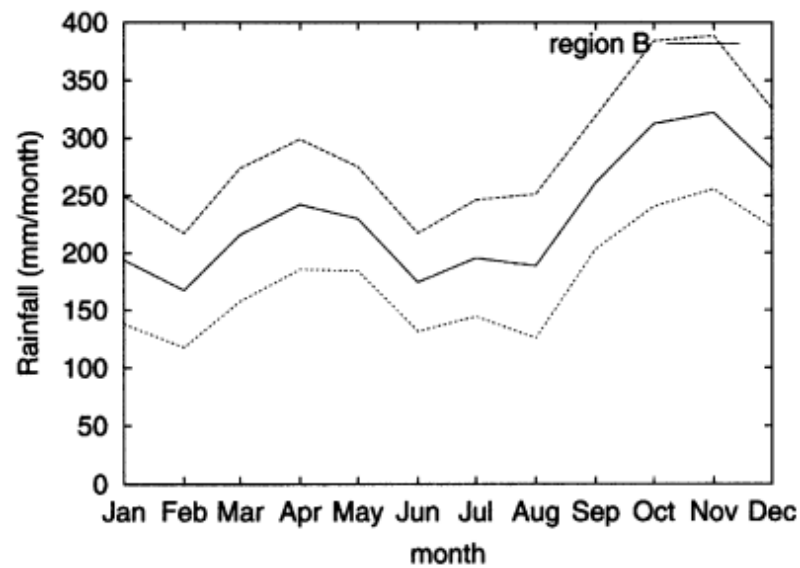
mei hingga September, pola hujan ini juga dapat kita sebut sebagai pola hujan Monsun, dimana terdapat perbedaan yang jelas antara periode musim hujan dan musim kemarau



Gambar 2. 2 Pola Curah Hujan Wilayah A

2. Wilayah B

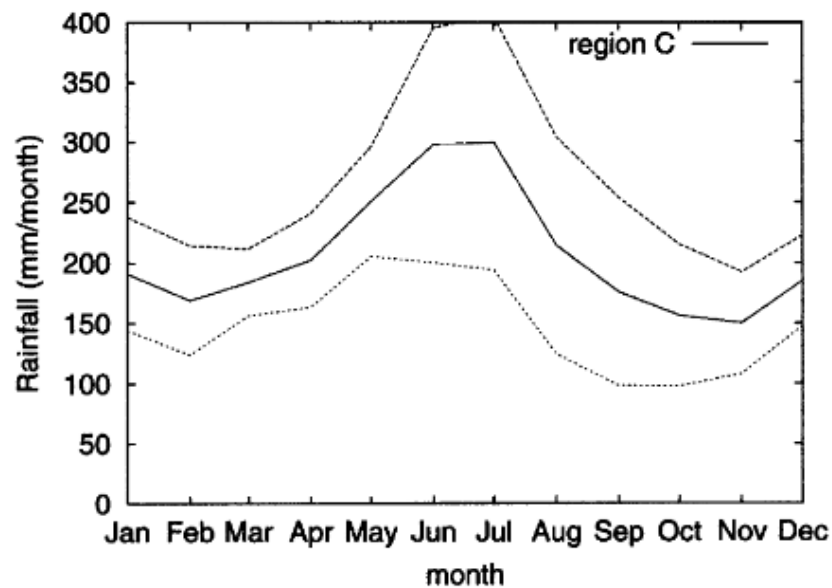
Terdapat 2 (dua) Puncak pada Pola Curah Hujan Wilayah B pada bulan oktober hingga November dan selanjutnya pada Maret hingga Mei, pola hujan ini dapat kita sebut dengan pola hujan Equatorial dengan ciri memiliki dua puncak musim hujan.



Gambar 2. 3 Pola Curah Hujan Wilayah B

3. Wilayah C

Pada wilayah C terdapat satu (satu) puncak curah hujan pada bulan Juni hingga Juli dan curah hujan terendah pada bulan november hingga Februari, pola hujan pada wilayah C memiliki nilai yang berkebalikan dengan Pola hujan yang berada pada wilayah A, pola hujan ini dapat kita sebut dengan pola hujan lokal.



Gambar 2. 4 Pola Curah Hujan Wilayah C

Dari gambar 2.1 dapat kita ketahui bahwa provinsi Lampung termasuk pada wilayah A yang kemudian dikelompokkan menjadi Zona Musim (ZOM), ZOM sendiri adalah daerah yang pola curah hujan rata-ratanya mempunyai suatu perbedaan yang jelas antara musim hujan dan musim kemarau.

2.3.4 Penakar Curah Hujan tipe Observasi

Pada penelitian ini alat yang digunakan untuk mengukur curah hujan yaitu menggunakan penakar hujan tipe observasi (obs), cara kerja penakar curah hujan tipe obs yaitu ketika air hujan jatuh ke permukaan bumi maka air akan tertampung pada corong penakar hujan yang selanjutnya akan tertampung pada tabung penampung dan di ukur pada gelas ukur dengan skala 0 – 25 mm, apabila jumlah curah hujan yang di tampung melebihi kapasitas gelas ukur, maka dilakukan pengukuran sampai beberapa kali sehingga curah hujan yang yg terdapat pada tabung dapat terukur semua sampai benar-benar habis, terdapat 4 bagian pada penakar curah hujan obs, yaitu :

1. Corong yang berbentuk lingkaran dengan luas 100 cm^2 dan dapat dilepas
2. Tabung penampung
3. Keran buka tutup
4. Peyangga kaki



Gambar 2. 5 Penakar Hujan tipe Observasi di stasiun Geofisika Lampung Utara

2.3.5 Suhu Udara

Suhu udara didefinisikan sebagai keadaan panas pada suatu benda atau luasan dalam waktu dan periode tertentu. Faktor utama penyebab suhu udara adalah sinar matahari pada suatu bidang atau area tertentu, sedangkan yang menyebabkan tinggi rendahnya suhu udara adalah sifat benda dan daerah tertentu seperti sifat pantul dan penyerap sinar matahari. Suhu udara permukaan dinyatakan dalam derajat Celsius ($^{\circ}\text{C}$). Suhu udara permukaan diukur dari ketinggian 1,20 - 1,25 meter di atas permukaan tanah dan diamati menggunakan Thermometer bola kering.

2.3.6 Penyinaran Matahari

Penyinaran matahari yang terpantau di stasiun geofisika Lampung Utara ditentukan oleh durasi penyinaran matahari selama periode waktu tertentu, yaitu dari pukul 06:00 WIB hingga 18:00 WIB. Alat yang digunakan untuk mengamati penyinaran matahari adalah Campbell Stokes (*Sun Shine Recorder*). Satuan untuk menentukan durasi lamanya penyinaran matahari adalah persentase (%) dan jam. Satuan persentase (%) digunakan untuk tujuan klimatologis dan satuan dalam jam untuk keperluan meteorologi.

2.3.7 Kelembaban Relatif

Kelembaban relatif adalah rasio massa uap air dalam satuan volume dengan massa uap air yang dibutuhkan untuk menjenuhkan volume udara pada suhu yang sama. Kelembaban relatif dinyatakan dalam persen (%). Alat yang digunakan untuk menentukan kelembaban relatif adalah Screen Psychrometer/Psychrometer sangkar tetap (Thermometer bola kering dan Thermometer bola basah) dan termohigrograf.

2.3.8 Tekanan Udara

Tekanan udara dapat dikatakan sebagai berat atmosfer pada suatu luasan bidang horizontal, barometer digital merupakan alat yang digunakan untuk mengukur tekanan udara, dimana tekanan dihitung dari permukaan stasiun (QFE) dan permukaan diatas permukaan laut (QFF), satuan tekanan udara dinyatakan dalam satuan millibar (mb) $1 \text{ millibar (mb)} = 1 \text{ hektopascal (HPa)}$.

2.3.9 Angin

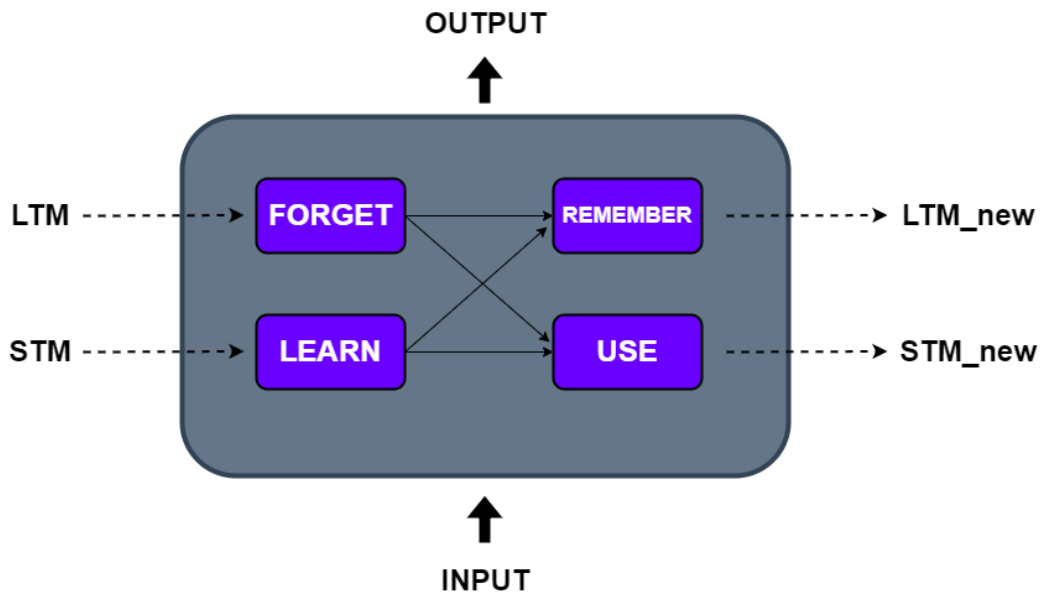
Angin adalah udara yang bergerak secara horizontal di permukaan bumi, Arah angin merupakan arah dari mana angin bergerak. Kecepatan angin adalah kuantitas vektor tiga dimensi dengan variasi pada skala kecil yang bergerak secara acak dalam ruang dan waktu yang menambah aliran pada skala besar secara teratur. Arah dan kecepatan angin permukaan diukur dari ketinggian 10 meter di atas permukaan tanah menggunakan anemometer

2.3.10 Long Short Term Memory (LSTM)

LSTM dirancang oleh Hochreiter dan Schmidhuber pada tahun 1997, Jaringan terdiri dari lapisan LSTM dengan proses berulang. Dalam modul LSTM terdapat beberapa perhitungan seperti penjumlahan, perkalian, penggabungan, penggandaan vektor, dan fungsi matematika seperti operasi tanh dan sigmoid, jika pada *Recurrent neural network* (RNN) menyambungkan neuron dan meneruskan ke depan, sedangkan LSTM meneruskan dan melihat neuron yang ada di belakangnya [20].

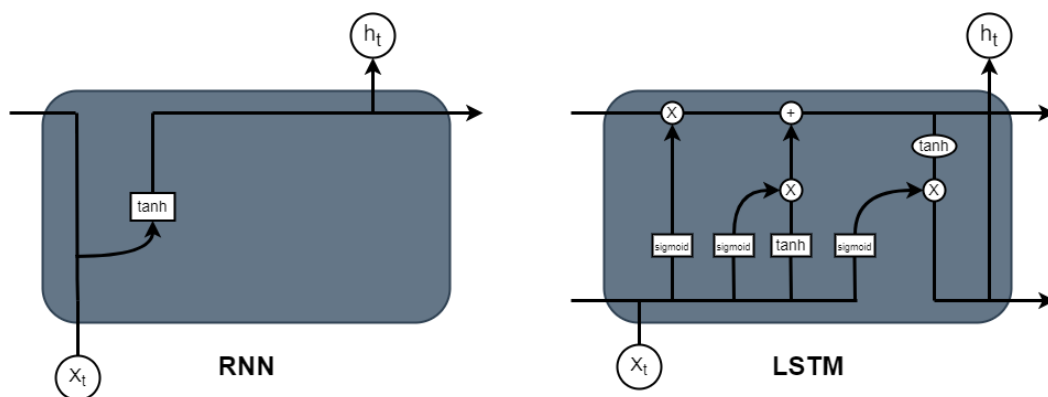
LSTM dibagi menjadi 4 gerbang (*gate*), *Forget*, *Learn*, *Remember* dan *Use*, Jika terdapat input dikombinasikan *short Term memory* (STM) maka akan menghasilkan *learn*, Jika terdapat input dan *long term memory* (LTM) maka akan menghasilkan *Forget* dimana *forget* adalah ingatan jangka panjang yang sudah tidak relevan.

1. *Forget* : $LTM * \text{forget factor}$
2. *Learn* : $STM + \text{input}$
3. *Remember* : $\text{not Forget} + \text{Learn} \rightarrow \text{New LTM}$
4. *Use* : $\text{Not Forget} + \text{Learn} \rightarrow \text{Prediction \& New STM}$



Gambar 2. 6 Arsitektur LSTM

LSTM di rancang untuk menangani persoalan pada RNN berupa *long-term dependency problem*, RNN dapat dapat kehilangan informasi penting di awal jika sequence cukup panjang, pada *backward propagation* RNN bias mengalami *vanishing gradient* dimana nilai gradient nya sangat kecil sehingga tidak berkontribusi pada *gate*, kasus ekstrim pada RNN menyebabkan kondisi stop training hal ini terjadi pada *sequence* yang lebih panjang 5 – 10 steps.



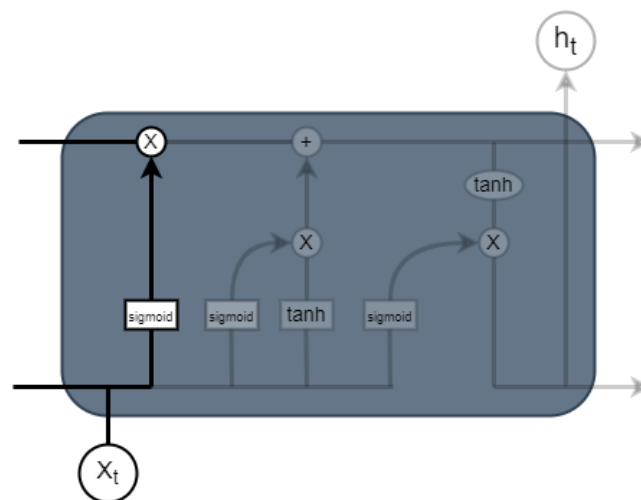
Gambar 2. 7 Perbandingan RNN dan LSTM

LSTM dirancang sebagai unit pengembangan dari RNN yang memiliki struktur yang sederhana sedangkan pada LSTM di usulkan untuk menambah *cell State* dan menambahkan 3 gate [21].

a. Struktur LSTM

Struktur LSTM terdiri dari 4 bagian, yaitu:

1. *Forget Gate*



Gambar 2. 8 Forget Gate

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f) \dots \dots \dots (1)$$

f_t : *Forget gate* aktivasi vektor

σ : Fungsi Sigmoid

x_t : *Input* vektor

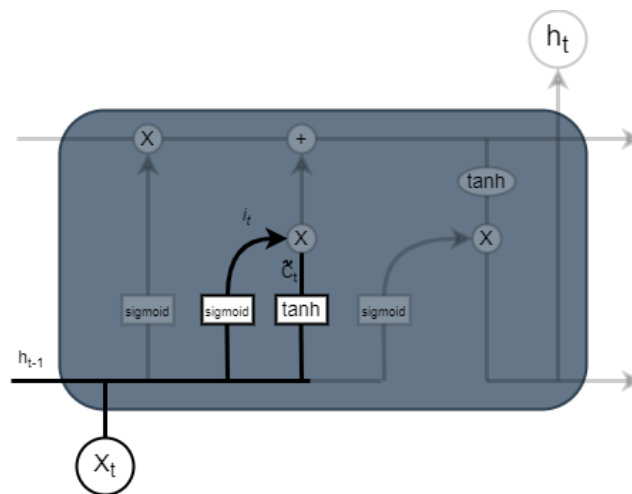
h_{t-1} : *Hidden state* vektor

U_f, W_f, b_f : Pembobotan matriks

Forget Gate menentukan informasi yang akan kita buang dari *cell state*, *forget gate* memproses h_{t-1} dan x_t jika dilihat dari persamannya di proses seperti pada RNN, terdapat matriks U , W dan b kemudian setelah dilakukan perkalian matriks

dimasukan pada fungsi aktifasi sigmoid sehingga f_t bernilai 0 – 1, jika nilai mendakti 1 maka informasi akan disimpan sedangkan jika mendekati 0 informasi nya akan di hapus

2. Input Gate



Gambar 2. 9 Input gate

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \dots \dots \dots (2)$$

$$\tilde{C}_t = \tanh(U_c x_t + W_c h_{t-1} + b_c) \dots \dots \dots (3)$$

i_t : *Input/Update* gate aktifasi vektor

\tilde{C}_t : *Cell input* aktifasi vektor

x_t : *Input* vektor

h_{t-1} : *Hidden state* vektor

σ : Fungsi Sigmoid

\tanh : Fungsi hyperbolic tangent

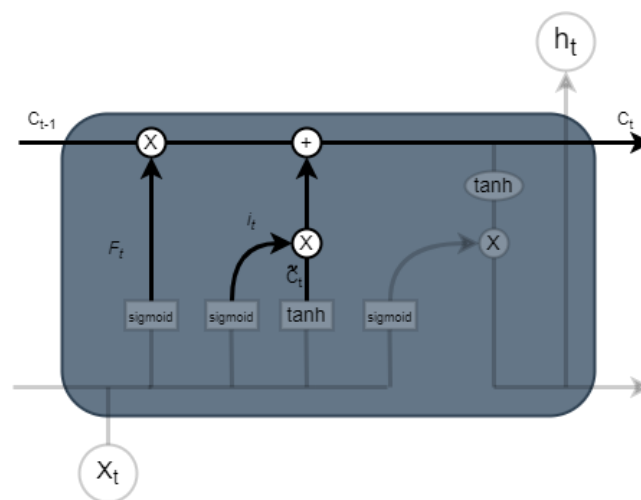
U_c, W_c, b_c : Pembobotan matriks

Penentuan informasi baru yang akan di simpan dalam *cell State*, input gate memiliki

2 operasi utama yaitu untuk menghitung nilai i_t dan nilai \tilde{C}_t , tetap memproses h_{t-1}

dan x_t tetap menggunakan matriks U , W dan b jadi setiap operasi memiliki matriks yang berbeda, kemudian i_t menggunakan sigmoid dan \tilde{C}_t menggunakan tanh, nilai i_t antara 0 – 1 sedangkan nilai \tilde{C}_t antara -1 – 1.

3. Cell State



Gambar 2. 10 Cell State

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \dots \dots \dots (4)$$

C_t : Cell state vektor

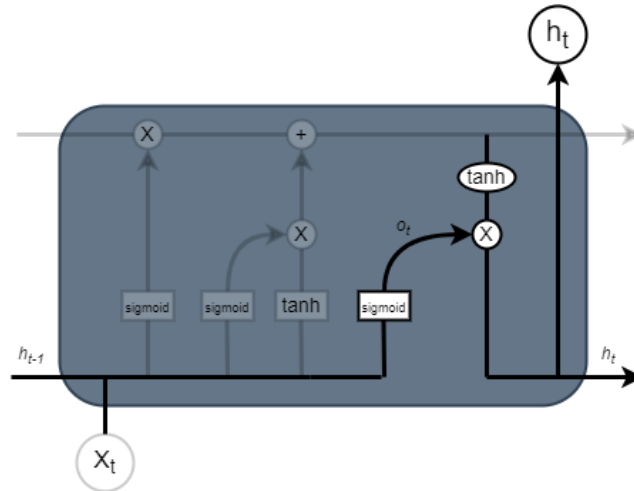
\tilde{C}_t : Cell input aktivasi vektor

f_t : Forget gate aktivasi vektor

i_t : Input/Update gate aktivasi vector

Merupakan unit tambahan yang dimiliki oleh LSTM dan tidak dimiliki pada unit RNN biasa, dapat kita lihat terdapat f_t dan i_t yang keduanya dihasilkan dari sigmoid yang artinya nilai keduanya antara 0 – 1 sedangkan \tilde{C}_t menggunakan fungsi tanh bernilai antara -1 – 1, fungsi dari f_t dan i_t sebagai bobot untuk menentukan informasi lama yang akan disimpan dan informasi baru yang akan di tambahkan.

4. Output Gate



Gambar 2. 11 Output gate

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \dots \dots \dots (5)$$

$$h_t = o_t * \tanh(C_t) \dots \dots \dots (6)$$

o_t : Output gate aktivasi vektor

h_t : Hidden state vektor

x_t : Input vektor

h_{t-1} : Hidden state vektor

σ : Fungsi Sigmoid

\tanh : Fungsi hyperbolic tangent

U_o, W_o, b_o : Pembobotan matriks

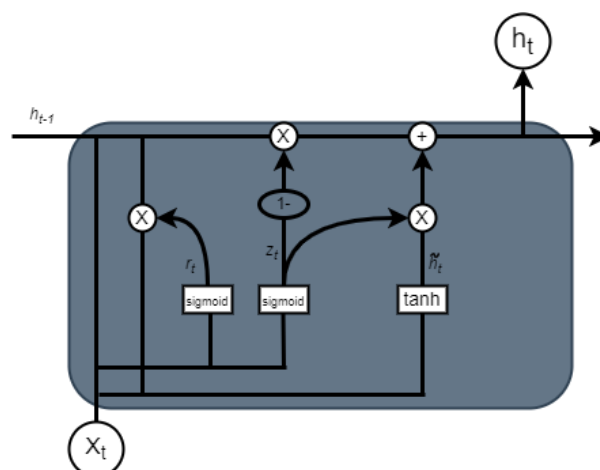
Output gate akan menentukan nilai dari h_t , kita sudah memiliki nilai dari C_t dan tetap menggunakan x_t dan h_{t-1} kemudian kita kalikan dengan matriks U_o , W_o dan b_o akan tetapi matrik ini khusus untuk output gate saja, kemudian hasil dimasukan ke fungsi σ sehingga o_t bernilai 0 – 1 kemudian nilai C_t dipakai untuk menentukan

nilai h_t akan tetapi sebelumnya digunakan fungsi tanh agar bernilai $-1 - 1$ lalu dikalikan dengan o_t sebagai bobotnya untuk menentukan nilai h_t [22].

2.3.11 Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) diperkenalkan pertama kali oleh Cho dkk pada tahun 2014, yang mempunyai tujuan untuk menyelesaikan *vanishing gradient problem* yang terdapat pada RNN, *vanishing gradient problem* sendiri adalah ketika nilai gradient mengecil bersamaan dengan waktu, jika nilai gradient tersebut mengecil maka tidak akan berkontribusi banyak pada *learning*.

GRU dapat dikatakan merupakan sebuah variasi dari LSTM karena keduanya banyak memiliki kesamaan, jadi GRU dan LSTM adalah peningkatan versi dari RNN, GRU menggunakan 2 (dua) gate yaitu *Update Gate* dan *Reset Gate* untuk memutuskan informasi apa yang dapat diteruskan pada output, kedua bagian tersebut dapat dilatih untuk menyimpan informasi terdahulu tanpa harus membuang atau menghapus informasi yg tidak berhubungan dengan prediksi [23].

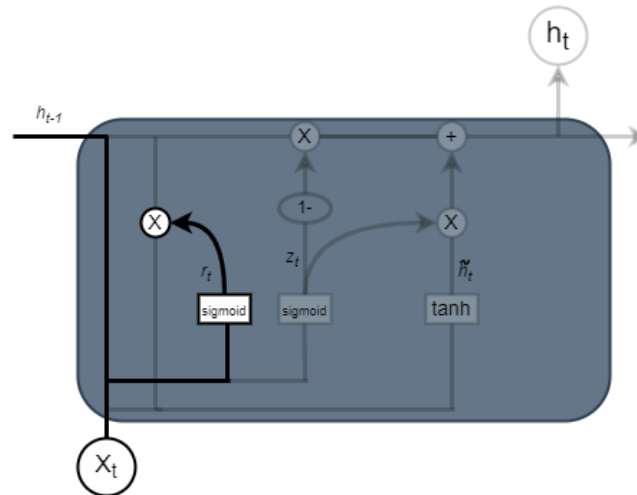


Gambar 2. 12 Diagram GRU

a. Struktur GRU

Struktur GRU terdiri dari 2 bagian, yaitu :

1. Reset Gate



Gambar 2. 13 Reset Gate

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r) \dots \dots \dots (7)$$

$$\tilde{h}_t = \tanh(W_h[r_t * h_{t-1}, x_t] + b_h) \dots \dots \dots (8)$$

r_t : Reset gate vektor

x_t : Input vektor

h_t : Output vektor

h_{t-1} : Hidden state vektor

σ : Fungsi Sigmoid

\tanh : Fungsi hyperbolic tangent

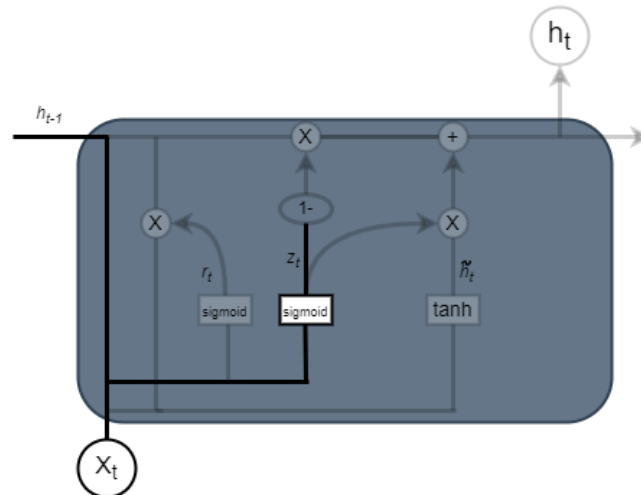
\tilde{h}_t : Kandidat aktivasi vektor

W_r, b_r, W_h, b_h : Pembobotan matriks

Reset gate terdiri dari *Hidden state* dan *input*, dengan bobot W_r dan b_r yang kemudian masuk ke sigmoid yang selanjutnya *reset gate* akan dikalikan dengan

elemen pada *hidden state* sebelumnya untuk memfilter berapa banyak *hidden state* sebelumnya yang akan di proses.

2. Update Gate



Gambar 2. 14 Update Gate

$$z_t = \sigma (W_z [h_{t-1}, x_t] + b_z) \dots \dots \dots (9)$$

$$\tilde{h}_t = \tanh(W_h [r_t * h_{t-1}, x_t] + b_h) \dots \dots \dots (10)$$

z_t : Update gate vektor

x_t : Input vektor

σ : Fungsi Sigmoid

\tanh : Fungsi hyperbolic tangent

h_t , : Output vektor

h_{t-1} : Hidden state vektor

\tilde{h}_t : Kandidat aktivasi vektor

W_z, b_z, W_h, b_h : Pembobotan matriks

Hampir menyerupai *reset gate*, *update gate* terdiri dari *hidden state* dan *input* dengan bobot W_z dan b_z lalu masuk ke sigmoid, terdapat *candidate activation vector* yang terdiri dari *hidden state time step* sebelumnya yang sudah di filter, input sekarang, bobot W_h dan b_h yang kemudian masuk pada fungsi tanh.

$$h_t = (1 - z_t) * h_t + z_t * \tilde{h}_t \dots \dots \dots (11)$$

h_t : *Output* vektor

\tilde{h}_t : Kandidat aktifasi vektor

z_t : *Update gate* vektor

Output akhir merupakan gabungan *hidden state* sebelumnya yang dikalikan 1-*update gate* dan *candidate activation vector* yang dikali dengan *update gate* dan kemudian dijumlahkan, sehingga *update gate* akan mengatur perbandingan antara *hidden state* sebelumnya dan *candidate activation vector* yang di hitung di *time step* sekarang.

2.3.12 Evaluasi

Evaluasi merupakan hal yang sangat penting untuk dilakukan untuk menentukan kinerja dan untuk menentukan kelebihan dan kekurangan suatu metode sehingga dapat ditingkatkan melalui penyesuaian dan modifikasi, evaluasi merupakan suatu metode menggunakan statistik dan data yang tersedia dengan cara menganalisis besarnya variasi dan menentukan kinerja suatu metode, diantara metode statistik yang sering digunakan untuk melakukan evaluasi adalah :

a. *Mean Square Error* (MSE)

Mean square error menghitung rata-rata kesalahan kuadrat antara nilai aktual dan prediksi, metode ini biasanya digunakan untuk mengecek estimasi nilai error dalam

suatu prediksi. Nilai MSE yang kecil atau nilai MSE yang mendekati nol berarti hasil prediksi sesuai dengan data aktual dan dapat digunakan untuk perhitungan prediksi periode mendatang. Metode MSE umumnya digunakan untuk mengevaluasi teknik pengukuran dengan menggunakan model regresi atau model prediksi seperti *moving average*, *weighted moving average* dan *trend analysis*. Cara menghitung *mean square error* (MSE) adalah dengan mengurangi nilai data aktual dengan data ramalan dan mengkuadratkan hasilnya kemudian menjumlahkannya secara keseluruhan dan membaginya dengan jumlah data yang ada.

$$MSE = \sum \frac{|\hat{y}_i - y_i|}{n} \dots\dots\dots(12)$$

\hat{y}_i : Nilai data prediksi

y_i : Nilai data sebenarnya

n : Banyaknya data

b. Mean Absolute Deviation (MAD)

Mean absolute deviation merupakan hitungan yang digunakan untuk menghitung rata-rata kesalahan absolut. MAD mengukur keakuratan prediksi dengan menyamakan ukuran kesalahan sehingga setiap prediksi memiliki nilai absolut untuk setiap kesalahan. MAD digunakan ketika ingin mengukur kesalahan perkiraan dalam unit pengukuran yang sama dengan data aslinya, MAD merupakan cara untuk menggambarkan variabilitas dari kumpulan data.

$$MAD = \frac{\sum |\hat{y}_i - y_i|}{n} \dots\dots\dots(13)$$

\hat{y}_i : Nilai data prediksi

y_i : Nilai data sebenarnya

n : Banyaknya data

c. Root Mean Square Error (RMSE)

RMSE digunakan untuk menghitung jumlah kesalahan dalam memprediksi suatu metode serta menghitung selisih antara nilai aktual dan nilai harapan dan membagi hasil total yang diperoleh dengan nilai hasil prediksi dan kemudian diakarkan.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \dots \dots \dots (14)$$

\hat{y}_i : Nilai data prediksi

y_i : Nilai data sebenarnya

n : Banyaknya data

d. Confusion Matrix

Confusion matrix merupakan hasil ringkasan kinerja dari proses klasifikasi yang berhubungan dengan data uji, dengan cara mengukur performanya sehingga kita dapat mengetahui seberapa baik sebuah model. Terdapat 4 kategori pada proses klasifikasi pada confusion matrix : True Positive (TP), False Positive (FP), True Negatif (TN) dan False Negative (FN) [24].

		NILAI SEBENARNYA	
		Positive (1)	Negative (0)
NILAI PREDIKSI	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. 15 Tabel confusion matrix

Confusion matrix memungkinkan kita untuk menghitung *performance metrics* untuk mengetahui keakuratan dari model yang kita gunakan, terdapat beberapa metode *performance metrics* yang sering digunakan yaitu : *Accuracy*, *Precision*, *Recall* dan *F1-Score*.

1. *Accuracy* : menghitung keakuratan model dalam melakukan klasifikasi

$$accuracy : \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(15)$$

2. *Precision* : menunjukkan tingkat keakuratan antara data yang diminta dengan hasil prediksi dari model, dengan kata lain merupakan rasio prediksi benar positif dibandingkan keseluruhan hasil prediksi positif.

$$Precision : \frac{TP}{TP+FP} \dots\dots\dots(16)$$

3. *Recall* : menggambarkan keberhasilan model dalam pencarian informasi. Jadi rasio prediksi benar positif terhadap semua prediksi benar-positif.

$$Recall : \frac{TP}{TP+FN} \dots\dots\dots(17)$$

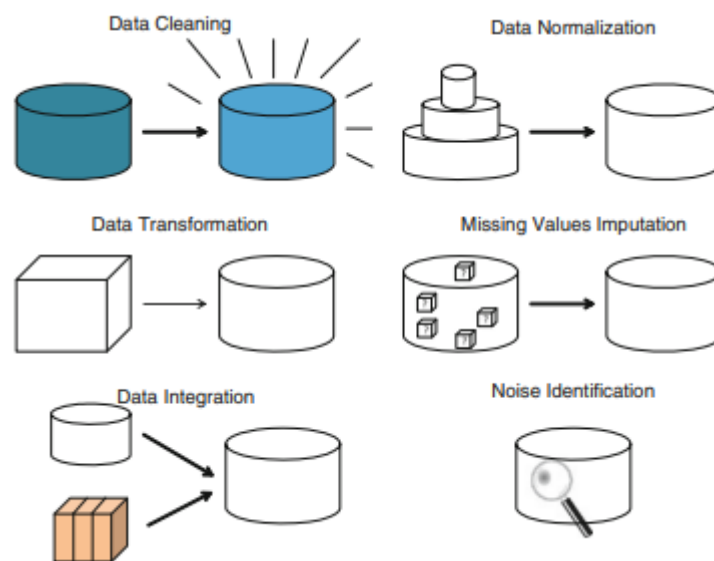
4. F1-score menunjukkan perbandingan nilai *precision* dan *recall* yang telah dibobotkan

$$F1 \text{ score} : \frac{2 * Recall * Precision}{Recall + Precision} \dots\dots\dots(18)$$

2.3.13 Preprocessing Data

Preprocessing merupakan bagian dari *data preparation* dan merupakan langkah yang wajib dilakukan untuk mengubah data yang tidak berguna menjadi data baru.

Preprocessing data adalah sebuah teknik yang dapat dilakukan pada dataset untuk menghilangkan data error, data yang tidak relevan, *missing data/value* serta *noise data*. Kegunaan preprocessing data adalah untuk mengubah data mentah sehingga dapat meningkatkan kualitas data yang akan di analisis.



Gambar 2. 16 Preprocessing Data

pada penelitian ini digunakan 2 (dua) metode *preprocessing* yaitu *data cleaning* dan *data transformation*

a. Data Cleaning

Data Cleaning atau *data cleansing* merupakan kegiatan mengisi data yang hilang, memfilter beberapa data yang salah dari kumpulan data dan mengurangi detail data yang tidak diperlukan, ada beberapa metode yang digunakan untuk melakukan *data cleaning* :

1. Mengabaikan struktur data
2. Mengisi nilai data yang hilang secara manual
3. Menggunakan *global constan* untuk mengisi data yang hilang
4. Menggunakan nilai *mean* dan *median*
5. Menggunakan nilai yang hilang dengan menggunakan regresi, decision tree dll.

b. Data Transformation

Data transformation digunakan untuk mengubah, menggabung dan menyamakan format data sehingga data yang akan dihasilkan lebih efisien dan lebih mudah dipahami sehingga dapat dibaca oleh algoritma, terdapat beberapa cara dalam melakukan *data transformation* :

1. *Smoothing* yaitu berfungsi untuk menghilangkan *noise* dari data
2. Membangun dan menambahkan atribut dari atribut yang sudah tersedia
3. Agregasi data dengan cara mengumpulkan dan menyajikan data dalam suatu format
4. Normalisasi dengan cara menskalakan atribut agar berada dalam rentang yang lebih kecil, contoh : $-1.0 - 1.0$ atau $0.0 - 1.0$

5. Discretization digunakan ketika data mentah yang berbentuk atribut angka di gantikan dengan sebuah interval label (contoh: 0-10, 11-20)
6. Menggeneralisasikan suatu atribut ke tingkatan yang lebih tinggi, contohnya adalah sebuah jalan menjadi suatu negara atau kota [25].

BAB III

METODE PENELITIAN