

Prediksi Pola Penyebaran Penyakit DBD di Kota Pagar Alam Menggunakan *Long Short Term Memory (LSTM)*

R. Ayu Syarifah Nurillah¹, Mahmud Imrona², Andry Alamsyah³,

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

³Fakultas Ekonomi Bisnis, Universitas Telkom, Bandung

¹ayusyarifah@students.telkomuniversity.ac.id, ²mahmudimrona@telkomuniversity.ac.id,

³andrya@telkomuniversity.ac.id,

Abstrak

DBD (Demam Berdarah *Dengue*) tumbuh secara cepat di seluruh dunia terlebih lagi di dunia yang memiliki iklim tropis seperti Indonesia, Kota Pagar Alam merupakan salah satu daerah yang masih terkena penyebaran penyakit DBD ini. Dalam beberapa dekade terakhir, variabel cuaca seperti suhu dan curah hujan telah dipelajari secara luas sebagai peringatan dini untuk mencegah penyakit menular yang sensitif terhadap iklim seperti *Dengue*. Salah satu cara untuk mencegah penyebaran penyakit DBD ialah membangun sistem yang dapat memprediksi penyebaran penyakit DBD. Penelitian ini menggunakan metode *Long Short Term Memory (LSTM)* dengan data historial penyakit DBD Rumah Sakit Umum Daerah Besemah dan data cuaca BMKG. Jaringan LSTM cocok untuk belajar dari pengalaman untuk mengklasifikasikan, proses dan memprediksi data time series ketika rentan waktunya sangat lama dari ukuran yang diketahui dengan kejadian yang penting. Ini adalah salah satu alasan utama mengapa LSTM menjadi alternatif dari RNNs dan Hidden Markov Model dan metode pembelajaran urutan lainnya dalam berbagai aplikasi. Penelitian ini melakukan percobaan dengan dataset yang tanpa di *preprocessing* dan di *preprocessing* terlebih dahulu dengan tiga model yang berbeda, yaitu model dengan 128 neuron, 64 neuron dan 32 neuron, yang menghasilkan konfigurasi terbaik pada dataset tanpa *preprocessing* dengan 32 neuron memiliki error 118.783 dan korelasi dari masing-masing daerahnya yaitu, Daerah A sebesar 0.594, Daerah B sebesar 0.420 serta Daerah C adalah sebesar 0.639. Serta pola penyebaran ini cenderung meningkat pada bulan oktober hingga februari, hal ini disebabkan karena curah hujan yang meningkat pada masa tersebut yang mengakibatkan perkembangbiakan nyamuk *aedes aegypti* juga meningkat.

Kata kunci: DBD, *long short term memory*, cuaca, regresi, *preprocessing*.

Abstract

Dengue (Dengue Hemorrhagic Fever) is growing rapidly throughout the world, especially in a world that has a tropical climate like Indonesia, Pagar Alam City is one of the areas that is still affected by the spread of this dengue disease. In recent decades, weather variables such as temperature and rainfall have been widely studied as early warnings to prevent climate-sensitive infectious diseases such as dengue. One way to prevent the spread of DHF is to build a system that can predict the spread of DHF. This study used the Long Short Term Memory (LSTM) method with historical data of dengue fever at the Besemah Regional General Hospital and BMKG weather data. LSTM networks are suitable for learning from experience to classify, process and predict time series data when the timeframe is too long from a known measure to important events. This is one of the main reasons why LSTM is an alternative to RNNs and Hidden Markov Model and other sequence learning methods in a wide variety of applications. This study conducted an experiment with a dataset that was not preprocessed and preprocessed first with three different models, namely a model with 128 neurons, 64 neurons and 32 neurons, which produced the best configuration in the dataset without preprocessing with 32 neurons having an error of 118,783 and a correlation of The respective regions were Region A amounting to 0.594, Region B amounting to 0.420 and Region C amounting to 0.639. And this spread pattern tends to increase in October to February, this is due to the increased rainfall during that period which resulted in the breeding of the *Aedes aegypti* mosquitoes also increasing.

Keywords: DHF, long short term memory, weather, regression, preprocessing.

1. Pendahuluan

DBD (Demam Berdarah *Dengue*) tumbuh secara cepat di seluruh dunia terlebih lagi di dunia yang memiliki iklim tropis seperti Indonesia, Kota Pagar Alam ialah salah satu daerah yang masih terkena penyebaran penyakit DBD ini. Menurut buku [1] total jumlah kasus penderita penyakit DBD pada tahun 2012-2014 adalah sebesar 209

kasus dan 1 angka kematian pada wilayah kerja Puskesmas Pengandonan dan menurut buku [2] jumlah kasus DBD pada tahun 2015 sebesar 77 kasus.

Demam Berdarah *Dengue* dapat menyerang siapa saja termasuk anak-anak. Tanda yang paling sering ialah berupa demam, nyeri pada otot dan nyeri sendi, yang disebabkan oleh virus *dengue* yang dibawa oleh nyamuk *aedes aegypti* [1]. Cuaca yang tak menentu serta tingginya curah hujan menyebabkan meningkatnya populasi nyamuk aedes [2]. Saat ini variabel cuaca seperti suhu dan curah hujan telah dipelajari secara luas sebagai peringatan dini untuk mencegah penyakit menular yang sensitif terhadap iklim seperti *Dengue* dan Malaria [3].

Sebelumnya telah ada penelitian yang dilakukan, jurnal [4] menyimpulkan akurasi dengan dua metode pembelajaran, yaitu akurasi terbaik menggunakan metode LSE didapat sebesar 65% dan akurasi terbaik yang didapatkan dengan metode *Backpropagation* adalah sebesar 70%. Dan jurnal [5] menghasilkan akurasi *testing* terbaik sebesar 75%.

Dalam tugas akhir ini, memprediksi pola penyebaran penyakit DBD menggunakan algoritma *Long Short Term Memory (LSTM)* dengan data historial penyakit DBD, cuaca serta suhu, yang diharapkan mampu memprediksi pola penyebaran penyakit DBD agar Dinas Kesehatan Kota Pagar Alam dapat mengambil keputusan lebih dulu untuk mengurangi perkembangan penyakit DBD. Prediksi pola penyebaran yang dimaksud oleh penulis dalam penelitiannya ini ialah bagaimana pola peningkatan atau penurunan jumlah pasien yang terserang DBD pada bulan-bulan tertentu dan apakah perubahan curah hujan serta suhu berperan penting dalam menentukan kenaikan jumlah pasien yang terserang DBD. *Long Short Term Memory (LSTM)* merupakan evolusi dari arsitektur *Recurrent Neural Network (RNN)*. Dalam perbandingan dengan RTRL, BPTT, *Recurrent Cascade-Correlation*, *Elman net*, dan *Neural Sequence Chunking*, LSTM menghasilkan lebih banyak proses sukses, dan belajar lebih cepat[6]. Jaringan LSTM cocok untuk belajar dari pengalaman untuk mengklasifikasikan, proses dan memprediksi data time series ketika rentan waktunya sangat lama dari ukuran yang diketahui dengan kejadian yang penting. Ini adalah salah satu alasan utama mengapa LSTM menjadi alternatif dari RNNs dan *Hidden Markov Model* dan metode pembelajaran urutan lainnya dalam berbagai aplikasi[7].

Topik dan Batasannya

Tugas akhir ini dilakukan untuk memprediksi pola penyebaran penyakit DBD (Demam Berdarah *Dengue*). Output dari tugas akhir ini merupakan prediksi pola penyebaran jumlah kasus dari masing-masing daerah yang terserang penyakit DBD dan angka error model sistem yang dibuat menggunakan metode *Long Short Term Memory (LSTM)*. Dalam tugas akhir ini terdapat dua rumusan masalah. Rumusan masalah yang pertama yaitu, bagaimana pola peningkatan atau penurunan kasus DBD. Rumusan masalah yang kedua yaitu, memprediksikan jumlah pasien yang terserang penyakit DBD pada tahun 2019 di daerah A, B, dan C. Lalu untuk batasan tugas akhir ini juga terdapat dua, yang pertama yaitu data yang digunakan merupakan data historial pasien dari Rumah Sakit Umum Daerah Besemah Kota Pagar Alam serta data cuaca yang didapat dari BMKG. Batasan masalah yang kedua yaitu, tugas akhir ini menggunakan metode *Long Short Term Memory (LSTM)*.

Tujuan

Tujuan dari tugas akhir ini adalah untuk mengetahui pola peningkatan atau penurunan kasus DBD dari masing-masing daerah menggunakan metode *Long Short Term Memory (LSTM)* dan mengetahui kinerja sistem dengan dilihat dari angka error yang didapat.

2. Studi Terkait

Sebelumnya terdapat beberapa penelitian yang memprediksikan masalah penyebaran penyakit demam berdarah ini dan hasilnya berbeda-beda. Pada tahun 2016, terdapat penelitian yang mempelajari tentang faktor penyebaran penyakit DBD. Faktor lingkungan umumnya terkait dengan penyebaran demam berdarah. Salah satu faktor lingkungan yang terkena penyakit DBD yaitu suhu. Dalam penelitian ini, hubungan antara suhu permukaan tanah, ketinggian, dan demam berdarah dipelajari. Oleh karena itu, penelitian ini menggunakan sistem informasi geografis dan citra penginderaan jauh dalam menerapkan analisis spasial. Penelitian ini menyimpulkan bahwa kasus demam berdarah didistribusikan sepenuhnya pada kisaran suhu dari 12°C hingga diatas 30°C [8].

Pada tahun 2017 terdapat penelitian yang menggunakan algoritma genetika. Penelitian ini menyimpulkan bahwa penerapan algoritma genetika mampu meningkatkan nilai akurasi, sebagai contoh dengan skenario pengujian 200 populasi dan 1000 generasi mampu menghasilkan akurasi sebesar 71,97%. Hasil yang didapat dari penelitian ini yaitu pola penyebaran penyakit demam berdarah di Kota Bandung dari tahun 2010 sampai 2014 sebesar 0% wilayah Kota Bandung bebas dari penyebaran penyakit DBD, 56,66% wilayah Kota Bandung siaga terhadap penyebaran penyakit DBD, 2% wilayah Kota Bandung bahaya penyebaran penyakit DBD dan 0% wilayah Kota Bandung endemik penyakit DBD [9].

Jurnal yang selanjutnya juga dilakukan di tahun 2017, penelitian ini memprediksikan demam berdarah menggunakan algoritma *Clustering K-Means* yang fokus terhadap empat tahap yaitu *preprocessing*, pemilihan atribut, pengelompokan dan prediksi demam berdarah. Menurutnya virus *dengue* ini dapat diprediksi dengan menerapkan berbagai teknik penambangan data. Selain itu, teknik pengelompokan adalah alat yang sangat baik untuk memvisualisasikan penyakit. Penelitian ini menyimpulkan hasil bahwa kisaran umur 0-15 tahun sebanyak 913 orang mengalami demam berdarah klasik/normal, kisaran umur 15-55 tahun sebanyak 650 orang mengalami demam berdarah *dengue* dengan *serotype* 1 sebanyak 327 orang dan *serotype* 2 sebanyak 323 orang, dan pada kisaran umur lebih dari sama dengan 55 tahun sebanyak 347 orang mengalami demam berdarah *dengue* dengan *serotype* 3 sebanyak 98 orang dan *serotype* 4 sebanyak 249 orang [10].

Selanjutnya pada tahun 2018 penelitian yang dilakukan menggunakan kernel dot dan radial pada algoritma *Support Vector Machine (SVM)* dengan data historial penyakit DBD dan cuaca dari tahun 2009 sampai 2016, pada kernel radial didapat akurasi testing sebesar 93% sedangkan pada kernel dot didapat rata-rata akurasi testing sebesar 62% [11].

Di tahun yang sama, yaitu tahun 2018 terdapat penelitian yang dilakukan untuk memprediksi penyebaran penyakit DBD menggunakan algoritma *Adaptive Neuro Fuzzy Inference System (ANFIS)* di Kabupaten Bandung dengan dua metode pembelajaran yaitu metode LSE yang menghasilkan akurasi testing sebesar 65% sedangkan pada metode pembelajaran *Backpropagation* menghasilkan akurasi testing sebesar 70%, penelitian ini menggunakan data cuaca yang berjumlah 6 atribut serta data historial penyakit DBD dan jumlah penduduk kabupaten Bandung periode 2009 sampai 2017 [4].

Dapat dilihat pada tahun 2018 terdapat dua metode yang digunakan berbeda dan akurasi yang didapat pun berbeda dikarenakan adanya perbedaan atribut pada data tersebut, penelitian yang menggunakan *Support Vector Machine (SVM)* hanya menggunakan data cuaca dan data historial penyakit DBD sedangkan pada penelitian yang menggunakan *Adaptive Neuro Fuzzy Inference System (ANFIS)* terdapat tambahan data yaitu data jumlah penduduk kabupaten Bandung.

2.1 DBD

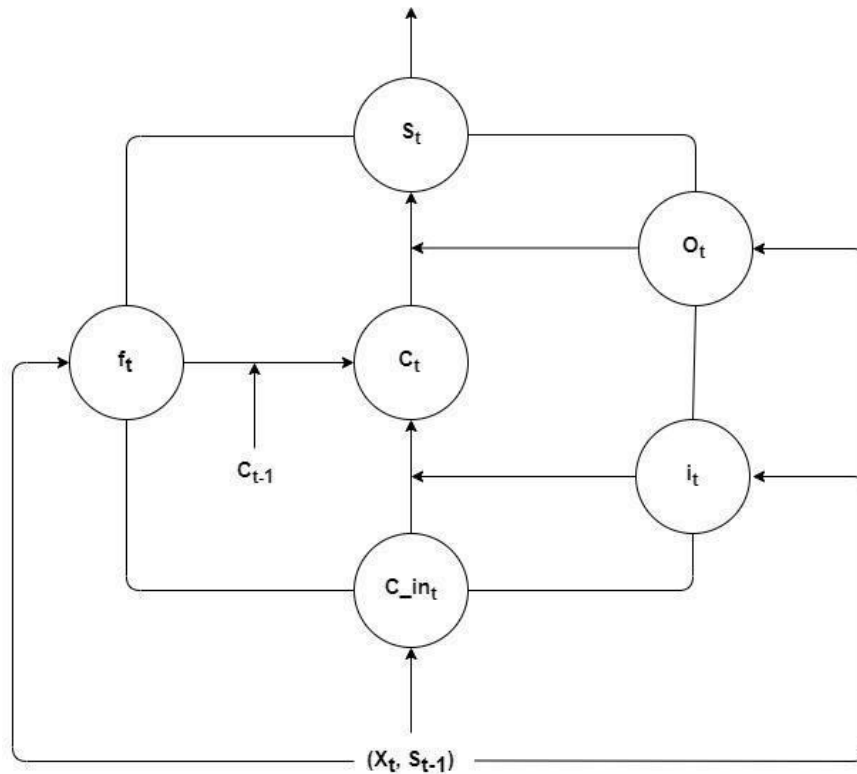
Demam Berdarah *Dengue* atau yang sering disingkat DBD banyak ditemukan di daerah tropis dan sub-tropis. Berdasarkan data dari seluruh dunia menunjukkan Asia menempati urutan pertama dalam jumlah penderita DBD setiap tahunnya. Terhitung sejak tahun 1968 hingga tahun 2009, *World Health Organization (WHO)* mencatat negara Indonesia sebagai negara dengan jumlah kasus DBD tertinggi di Asia Tenggara [12].

Demam Berdarah *Dengue* adalah infeksi yang disebabkan oleh virus *dengue*, *dengue* merupakan penyakit yang ditularkan dari nyamuk *Aedes SPP*, nyamuk yang paling cepat berkembang di dunia ini dan telah menyebabkan hampir 390 juta orang terinfeksi setiap tahunnya. Untuk Indonesia dengan iklim tropis sangat cocok untuk pertumbuhan hewan ataupun tumbuhan dan baik bagi tempat berkembangnya berbagai penyakit, terutama penyakit yang dibawa oleh vektor, yakni organisme penyebar agen patogen dari inang ke inang, seperti nyamuk yang banyak menularkan penyakit [13].

Penyebaran DBD dapat dikurangi melalui perilaku hidup sehat dan lingkungan yang bersih, melakukan *fogging* serta upaya-upaya lain yang dapat mencegah terjadinya perkembangan penyakit DBD. Selain perilaku hidup sehat dan lingkungan yang bersih, cuaca juga menjadi faktor penyebaran penyakit DBD disuatu daerah. Cuaca juga berperan dalam meningkatnya populasi nyamuk aedes.

2.2 Long Short Term Memory

Long Short Term Memory (LSTM) diciptakan oleh Sepp Hochreiter dan Jürgen Schmidhuber pada tahun 1997. *Long Short Term Memory (LSTM)* dapat digunakan sebagai unit *nonlinear* yang kompleks untuk membangun jaringan saraf yang lebih dalam, yang dapat mencerminkan efek memori jangka panjang dan memiliki kemampuan pembelajaran yang mendalam [14].

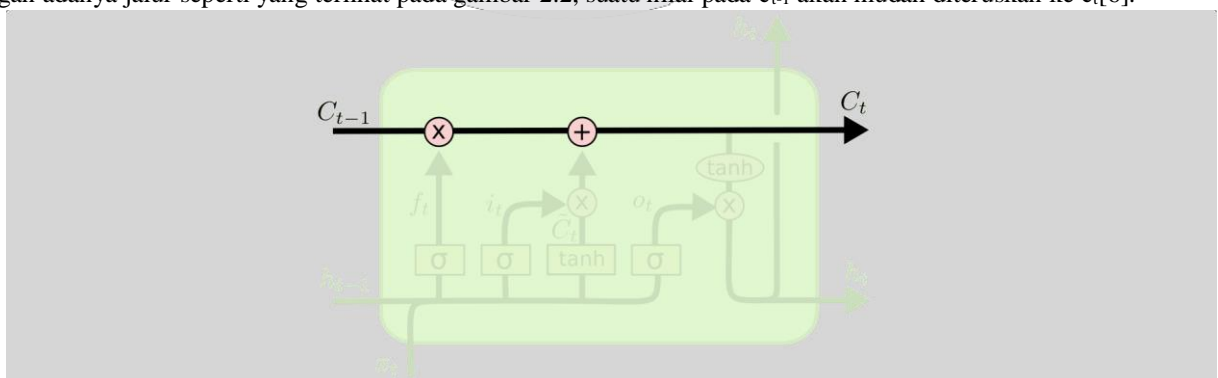


Gambar 2. 1 Struktur dasar LSTM [14]

Jaringan *Long Short Term Memory (LSTM)* terdiri dari lapisan *input*, *output* dan beberapa lapisan tersembunyi. Lapisan tersembunyi yang terdiri dari sel memori, struktur dasar yang ditunjukkan pada Gambar 2.1. Satu sel terdiri dari tiga *gates (input, forget, output)*, dan unit koneksi berulang. *Input* ke unit ini adalah x_t , *input* saat ini pada langkah t , dan s_t yaitu status tersembunyi saat ini. *Output* saat ini adalah o_t , dan c_t adalah memori internal unit [14]. *Forget gate* ialah langkah awal yang berfungsi untuk memutuskan informasi apa yang akan disimpan dan dibuang dari c_{t-1} . Selanjutnya, langkah untuk memutuskan informasi baru mana yang akan digunakan di c_t , langkah ini memiliki dua bagian yaitu *input gate* dan aktivasi *tanh*. Setelah itu, proses memperbarui informasi lama c_{t-1} ke informasi baru c_t . *Output gate* merupakan proses yang memutuskan informasi apa yang akan dihasilkan dan h_t yang menggunakan fungsi aktivasi *tanh*.

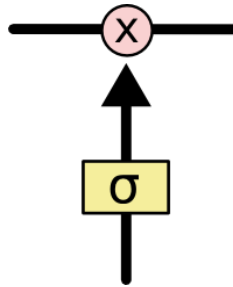
Long Short Term Memory (LSTM) merupakan evolusi dari arsitektur *Recurrent Neural Network (RNN)*. *Recurrent Neural Network (RNN)* tidak dapat untuk belajar menghubungkan informasi karena memori lama yang tersimpan akan semakin tidak berguna dengan seiringnya waktu berjalan karena tertimpa atau tergantikan dengan memori baru, permasalahan ini ditemukan oleh Bengio, et al. (1994). Berbeda dengan *Recurrent Neural Network (RNN)*, *Long Short Term Memory (LSTM)* tidak memiliki kekurangan tersebut karena *Long Short Term Memory (LSTM)* dapat mengatur memori pada setiap masukannya dengan menggunakan *memory cells* dan *gate units* [15].

Ide kunci dari model LSTM ini adalah jalur yang menghubungkan c_{t-1} ke c_t di bagian atas modul LSTM, dengan adanya jalur seperti yang terlihat pada gambar 2.2, suatu nilai pada c_{t-1} akan mudah diteruskan ke c_t [6].



Gambar 2. 2 Jalur yang menghubungkan c_t dengan c_{t-1} [6]

Ide kunci lain ialah adanya sigmoid *gate* (gerbang sigmoid) yang berfungsi untuk mengatur berapa banyak informasi bisa lewat. Gerbang sigmoid ini terdiri dari lapisan jaringan saraf sigmoid dan operasi perkalian elemen[6].



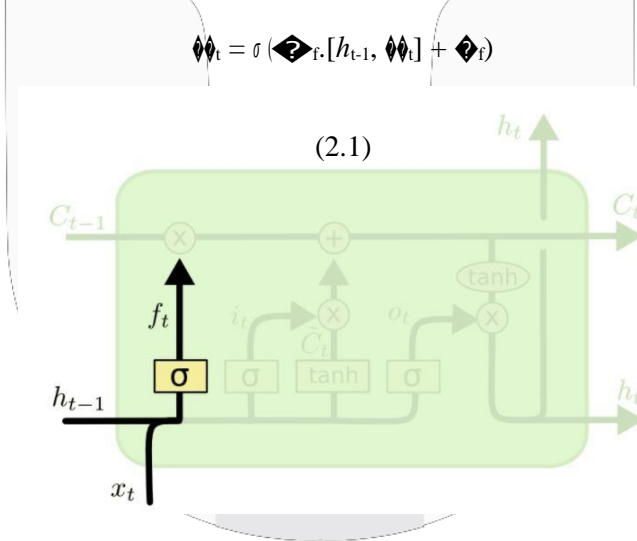
Gambar 2. 3 Gerbang sigmoid[6]

Keluaran dari sigmoid *gate* (gerbang sigmoid) ini merupakan angka 0 dan 1, jika 0 berarti informasi tersebut diblok atau tidak diperbolehkan masuk sedangkan angka 1 yang berarti informasi diperbolehkan masuk. Lalu keluaran dari sigmoid *gate* (gerbang sigmoid) ini akan dikalikan dengan suatu nilai untuk mengontrol berapa banyak informasi tersebut dipakai[6].

2.2.1 Memory Cells dan Gate Units

Seperti yang telah dijelaskan pada sub bab sebelumnya, dimana *Long Short-Term Memory (LSTM)* mempunyai *memory cells* dan *gate units* pada setiap *neurons*nya yang berfungsi untuk mengatur memori pada setiap *neurons*. Terdapat empat proses fungsi aktivasi pada setiap masukan pada *neurons* yang selanjutnya disebut sebagai *gate units*. *Gate units* tersebut ialah *forget gates*, *input gates*, *cell gates*, dan *output gates* [15].

Pada *forget gates* informasi pada setiap data masukan akan diolah dan dipilih data mana yang akan disimpan atau dibuang pada *memory cells*. Pada *forget gates* ini menggunakan fungsi aktivasi sigmoid, dimana hasil keluarannya 0 dan 1. Jika keluarannya 1, maka semua data akan disimpan dan sebaliknya. Dengan rumus sebagai berikut [15]:



Gambar 2. 4 Forget gate[6]

Contohnya model bahasa yang mencoba memprediksi kata selanjutnya berdasarkan dari semua kata sebelumnya. Dalam permasalahan tersebut, suatu elemen dalam c_{t-1} mungkin menyimpan informasi gender saat ini, sehingga kata ganti yang benar dapat digunakan. Ketika terdapat subjek baru, maka elemen lama dalam c_{t-1} dapat dilupakan[6].

Pada *input gates* terdapat dua *gates* yang akan dilaksanakan, pertama akan diputuskan nilai mana yang akan diperbarui menggunakan fungsi aktivasi sigmoid. Selanjutnya fungsi aktivasi tanh akan membuat vektor nilai baru yang akan disimpan pada *memory cell*. Dengan rumus sebagai berikut [15]:

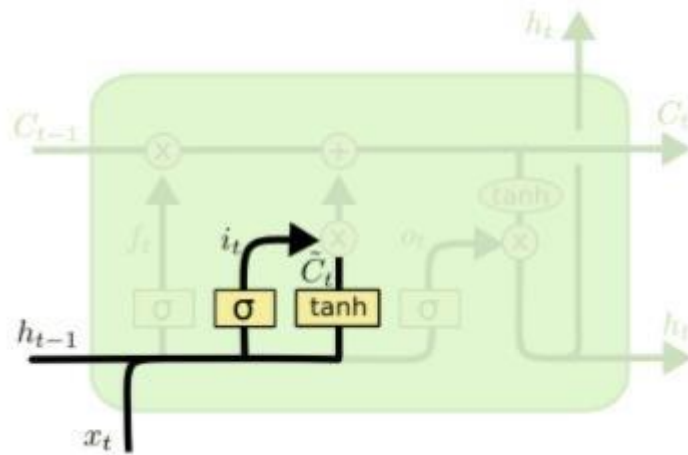
$$\hat{h}_t = \sigma(\hat{w}_t \cdot [h_{t-1}, \hat{h}_t] + \hat{b}_t)$$

(2.2)

$$\check{c}_t = \tanh(\check{w}_t \cdot [h_{t-1}, \hat{h}_t] + \check{b}_t)$$

(2.3)

Dalam contoh model bahasa yang dilakukan, ingin menambahkan gender subjek baru ke status sel, untuk mengganti yang lama yang telah dilupakan[6].

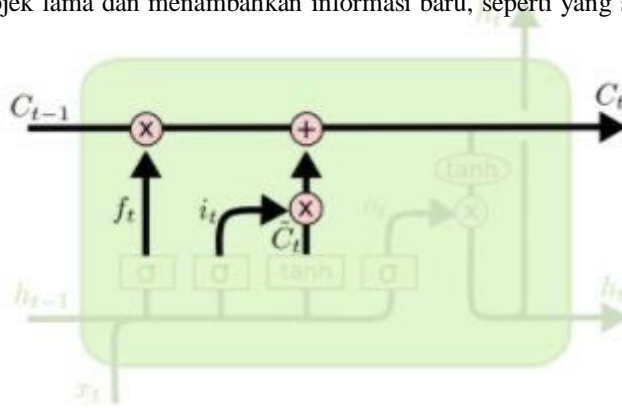


Gambar 2. 5 Input gate[6]

Pada *cell gates*, nilai pada *memory cell* sebelumnya akan diganti dengan nilai *memory cell* yang baru. Nilai ini didapatkan dari menggabungkan nilai yang ada pada *forget gate* dan *input gate*. Dengan rumus sebagai berikut [15]:

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t \tag{2.4}$$

Dalam contoh kasus model bahasa tadi, bagian inilah yang benar-benar melepaskan atau menggantikan informasi tentang gender subjek lama dan menambahkan informasi baru, seperti yang sudah putuskan di langkah sebelumnya[6].



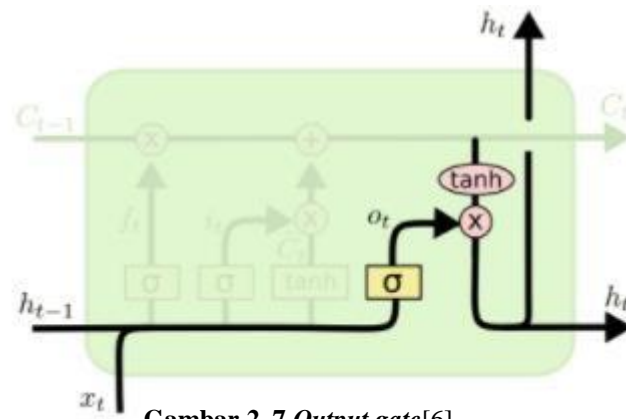
Gambar 2. 6 Cell gate[6]

Pada *output gates* terdapat dua *gates* yang akan dilaksanakan, pertama akan diputuskan nilai pada bagian *memory cell* yang akan dikeluarkan menggunakan fungsi aktivasi sigmoid. Selanjutnya akan ditempatkan nilai pada *memory cell* menggunakan fungsi aktivasi tanh. Terakhir kedua *gates* tersebut dikalikan sehingga menghasilkan nilai yang akan dikeluarkan. Dengan rumus sebagai berikut [15]:

$$o_t = \sigma(w_o \cdot [h_{t-1}, \check{C}_t] + b_o) \tag{2.5}$$

$$h_t = o_t * \tanh(C_t) \tag{2.6}$$

Untuk contoh kasus model bahasa tadi, karena hanya melihat subjek, kemungkinan ingin mengeluarkan informasi yang relevan dengan kata kerja, jika itu yang akan data berikutnya. Misalnya menampilkan apakah subjek tunggal atau jamak, sehingga diketahui bentuk kata kerja apa yang dikonstruksikan jika itu yang terjadi selanjutnya[6].



Gambar 2. 7 Output gate[6]

Keterangan:

f_t : Forget gate

i_t : Input gate

\tilde{c}_t : Vektor nilai kandidat baru

\diamond_t : Cell state baru

\diamond_t : Output gate

$\diamond\diamond$: Gerbang sigmoid

\diamond_{t-1} : Cell state lama

W_f : Bias forget

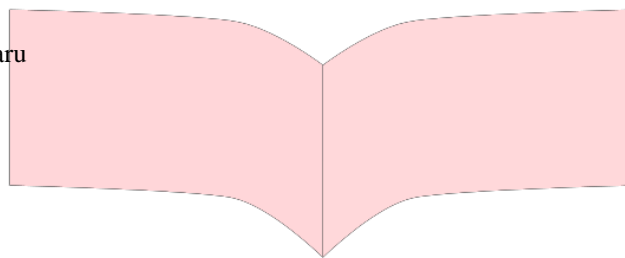
W_i : Bias input

W_c : Bias memory cell

W_o : Bias output

x_t : Input vector

h_t : Hidden state vector



2.3 Regresi

Istilah regresi pertama kali digunakan oleh Sir Francis Galton pada penelitiannya yang berhubungan dengan tinggi manusia, yaitu antara tinggi badan anak dan tinggi badan orang tuanya. Regresi merupakan alat ukur yang berfungsi untuk membantu memperkirakan nilai suatu variabel yang tidak diketahui dari satu atau lebih variabel lainnya dan mengetahui hubungan antar variabel tersebut. Regresi biasanya digunakan untuk prediksi atau ramalan[16].

2.4 Cuaca

Demam berdarah merupakan penyakit yang ditularkan oleh vektor nyamuk. Perkembangan vektor penyakit dapat dipengaruhi oleh perubahan iklim melalui berbagai cara yaitu, pertama unsur cuaca mempengaruhi metabolisme, pertumbuhan, perkembangan dan populasi nyamuk tersebut, kedua curah hujan dengan penyinaran yang relatif panjang turut mempengaruhi habitat perindukan nyamuk[17].

Demam berdarah *dengue* atau DBD adalah salah satu penyakit yang sensitif terhadap perubahan cuaca. DBD diperkirakan akan menonjol pada tahun 2080, sekitar 6 miliar orang akan berisiko tertular penyakit DBD sebagai konsekuensi dari perubahan iklim dibandingkan dengan 3-5 miliar orang jika iklim tetap tidak berubah [17].

3. Sistem yang Dibangun

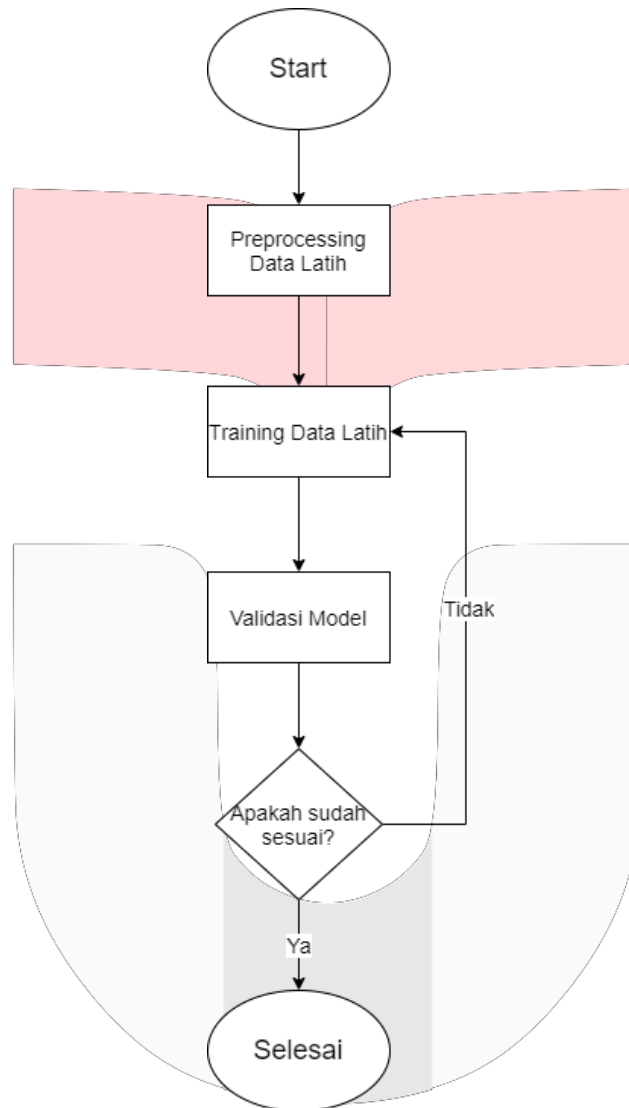
3.1 Data

Data yang penulis gunakan pada penelitian ini yaitu data historial penyakit DBD Rumah Sakit Umum Daerah Besemah yang berjumlah 9.061 data dengan rentang tahun 2015-2019 dan data suhu, kelembaban udara serta curah hujan yang didapatkan dari BMKG. Dataset tersebut terbagi menjadi datatrain yaitu dari tahun 2015 sampai tahun 2018, dan data validasi pada tahun 2019. Dataset terdiri dari 12 atribut, yaitu jumlah perempuan, jumlah laki-laki, jumlah bayi, jumlah anak, jumlah remaja, jumlah dewasa, jumlah lansia, curah hujan di daerah A, curah hujan di daerah B, curah hujan di daerah C, suhu, dan kelembaban serta target yang terdiri dari jumlah di daerah A, jumlah

di daerah B, jumlah di daerah C. Atribut inilah yang akan mempengaruhi hasil dari regresinya. Dan daerah A, B, dan C yang dimaksud disini ialah nama daerah yang telah disamakan oleh penulis.

Data mentah diproses dengan cara mengelompokan data menjadi data bulanan, selanjutnya data tersebut dihitung jumlah bayi (0-1 tahun), anak-anak (2-10 tahun), remaja (11 - 19 tahun), dewasa (20 - 60 tahun), lansia (>60 tahun), jumlah yang terjangkit di daerah A, daerah B, dan daerah C. Selanjutnya data curah hujan dipisahkan untuk daerah A, B, dan C. Terakhir data kelembaban dipisahkan tiap bulannya. Semua data tersebut kecuali jumlah yang terjangkit DBD di daerah A, B, dan C digunakan untuk melatih model yang dibuat.

3.2 Rancangan Sistem



Gambar 3. 1 Rancangan sistem

Tahapan pertama pada sistem ini yaitu melakukan *cleaning data* yang berfungsi untuk membersihkan data dan merapihkan data yang akan digunakan. Setelah melakukan *cleaning data*, data akan masuk ke proses *preprocessing* yang berfungsi meminimalisir error. *preprocessing* pada data train ini yaitu dengan mengubah data aktual menjadi nilai dengan range interval [0,1]. Teknik *preprocessing* yang penulis gunakan yaitu dengan teknik *min-max scaling* dengan rumus sebagai berikut[15]:

$$X' = \frac{(x - \min_x)}{(\max_x - \min_x)} \tag{3.1}$$

Keterangan:

X: Data yang akan di-*preprocessing*

X': Data setelah di-*preprocessing*

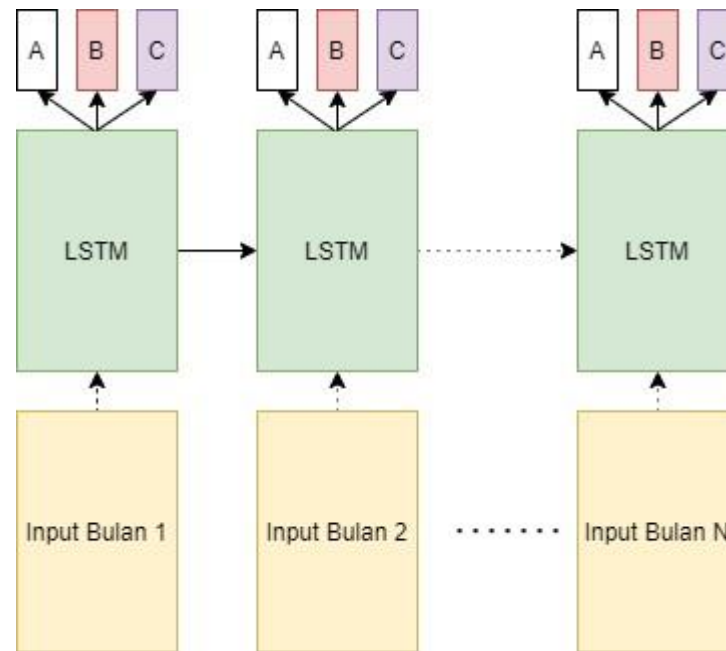
min_x : Nilai minimum dari keseluruhan data

max_x : Nilai maksimum dari keseluruhan data

Setelah dilakukan *preprocessing*, data train yang sudah siap digunakan akan dilakukan training data train yaitu data train akan dilatih dengan algoritma LSTM agar menghasilkan model program yang sesuai dan dapat menghasilkan akurasi terbaik ketika akan divalidasi pada proses berikutnya. Selanjutnya yaitu merupakan proses validasi model dimana pada proses inilah yang menentukan apakah model yang dibangun sudah sesuai dan sudah mendapatkan hasil akurasi terbaik, jika model yang dibangun belum sesuai maka program akan kembali melakukan proses training data train. Proses validasi model ini ialah proses terakhir dari program ini, proses yang mengukur performa dari metode program yang digunakan.

3.3 Desain Model

Model yang dibangun menggunakan satu buah layer LSTM dan satu buah layer Dense sebagai layer outputnya.



Gambar 3. 2 Arsitektur LSTM

Pada layer LSTM menerima input sebesar 12 fitur, dan memiliki *hidden layer* sebanyak 128 neuron. Pada layer output terdapat 3 neuron dengan fungsi aktivasi-linear yang merepresentasikan jumlah pasien tiap lokasi. Pada proses training menggunakan Adam[19] sebagai *optimizer*-nya, dan Mean Squared Error[15] sebagai *loss function*-nya dikarenakan outputnya berupa bilangan real.

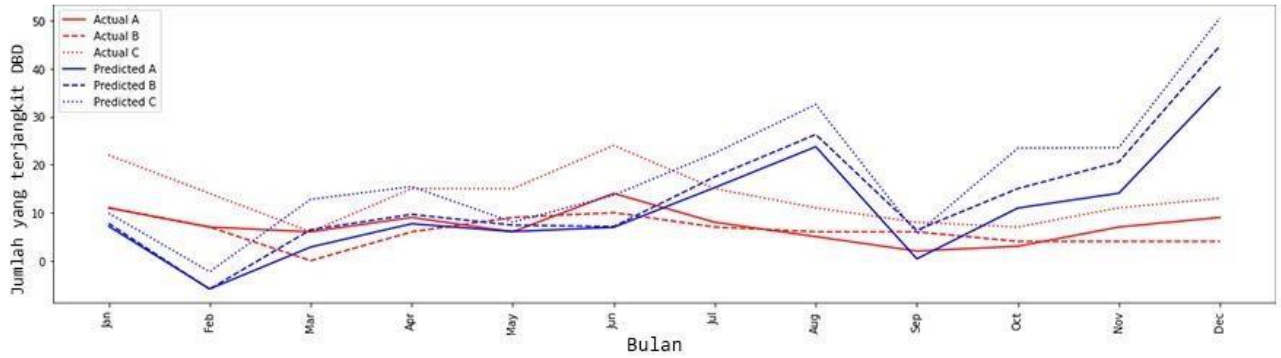
Berdasarkan arsitektur yang penulis gunakan, inputan yang digunakan berupa data yang sudah digabungkan menjadi data bulanan. Data tiap bulan tersebut akan digunakan sebagai input setiap timestep t , lalu setiap timestep akan menghasilkan tiga buah output, yaitu jumlah pasien yang terjangkit DBD di tiga daerah, yaitu daerah A, B, dan C.

4. Evaluasi

Pada bagian ini penulis melakukan dua skenario pengujian. Penulis melakukan pengujian menggunakan data historial penyakit DBD dan data BMKG. Dataset tersebut terbagi menjadi beberapa bagian yaitu, data train dan data validasi. Data train digunakan untuk melatih atau mengembangkan model program, data validasi digunakan untuk menguji model program. Skenario yang dilakukan penulis yaitu yang pertama, dataset tanpa *preprocessing* dan yang kedua yaitu dataset dengan *preprocessing*. Setiap skenario akan diujikan dengan 3 model yang berbeda, yaitu model LSTM dengan 32 neuron, 64 neuron, dan 128 neuron.

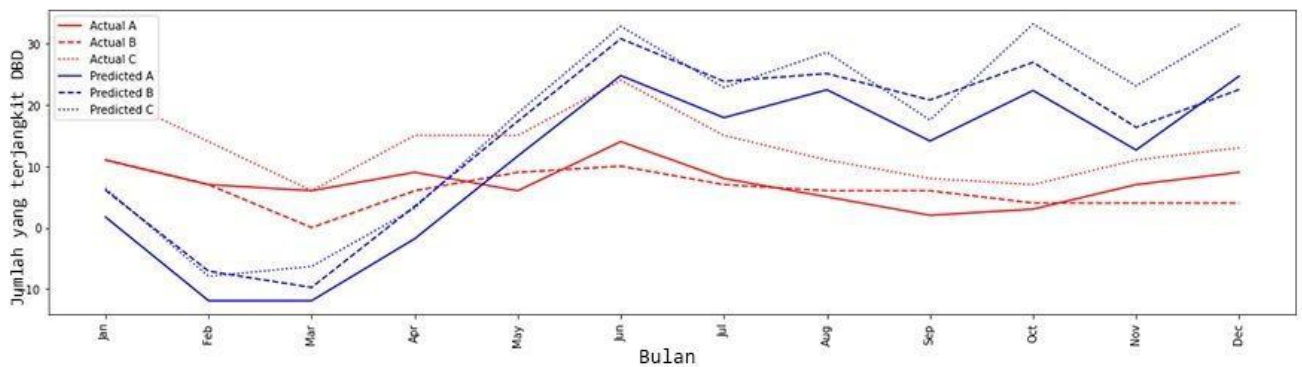
4.1 Hasil Pengujian tanpa *Preprocessing*

Pada eksperimen ini, skenario yang dilakukan adalah tanpa mengaplikasikan *preprocessing* pada dataset, dengan pembagian data train dari tahun 2015 sampai 2018 data dan data validasi pada tahun 2019. Untuk konfigurasi 128 neuron didapatkan hasil sebagai berikut:



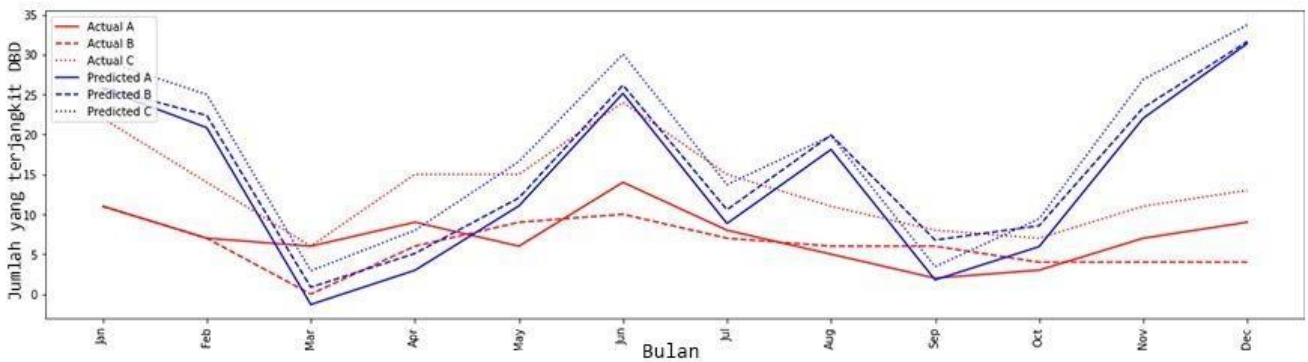
Gambar 4. 1 Grafik hasil percobaan 128 neuron dan dataset tanpa *preprocessing*

Untuk konfigurasi 64 neuron didapatkan hasil grafik sebagai berikut:



Gambar 4. 2 Grafik hasil percobaan 64 neuron dan dataset tanpa *preprocessing*

Untuk konfigurasi 32 neuron didapatkan hasil sebagai berikut:



Gambar 4. 3 Grafik hasil percobaan 32 neuron dan dataset tanpa *preprocessing*

Dari grafik diatas, didapatkan bahwa kenaikan yang terjadi pada data validasi juga dialami oleh data hasil prediksi, walaupun jumlah kenaikannya terjadi pada bulan selanjutnya dan terkadang terlalu tinggi dari yang seharusnya. begitu pula jika terjadi penurunan pada data validasi, terjadi penurunan juga pada data prediksi.

Dari hasil percobaan ini didapatkan hasil error sebagai berikut:

Tabel 1. Hasil error untuk dataset tanpa preprocessing

Konfigurasi	Error
128 neuron	238.13
64 neuron	168.53
32 neuron	118.783

Dari hasil diatas, didapatkan hasil korelasi sebagai berikut:

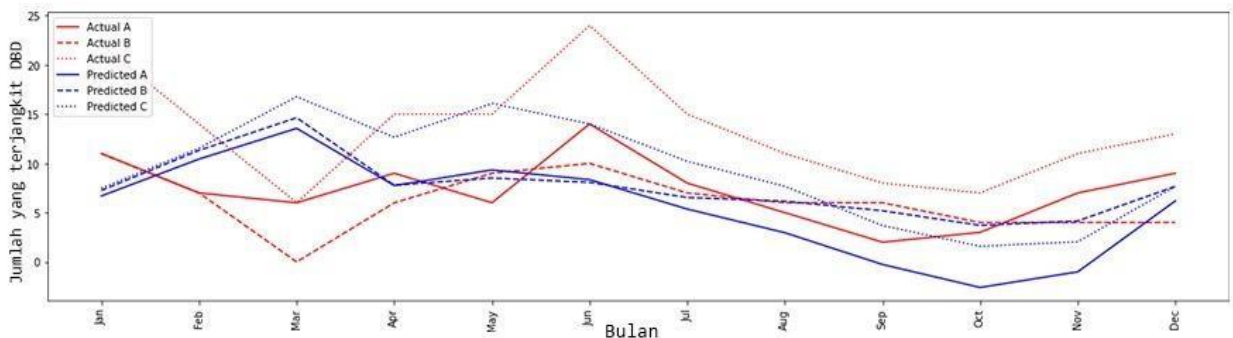
Tabel 2. Korelasi dari masing-masing daerah untuk dataset tanpa preprocessing

Daerah	Korelasi 128-neuron	Korelasi 64 neuron	Korelasi 32 neuron
Daerah A	0.116	0.015	0.594
Daerah B	-0.295	0.269	0.420
Daerah C	-0.156	0.076	0.639

Korelasi dengan nilai mendekati 1 berarti pola antara hasil prediksi dan data sebenarnya searah sedangkan jika nilai korelasinya mendekati -1 berarti pola antara hasil prediksi dan data sebenarnya berbanding terbalik. dalam pengujian ini hasil korelasi dari Daerah A data prediksi dan data sebenarnya memiliki pola searah sedangkan untuk Daerah B dan Daerah C memiliki pola berbanding terbalik pada 128 neuron. Sedangkan pada 64 dan 32 neuron data prediksi dan data sebenarnya pada masing-masing daerah searah.

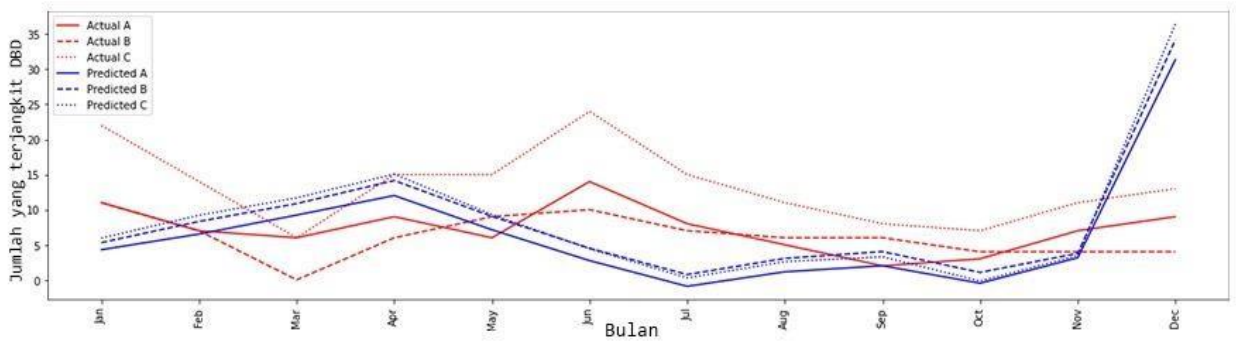
4.2 Hasil Pengujian dengan Preprocessing

Pada eksperimen ini, skenario yang dilakukan adalah dataset dengan menggunakan preprocessing min-max scaling, dan dengan pembagian data train dari tahun 2015 sampai 2018 data dan data validasi pada tahun 2019 sehingga didapatkan sebagai berikut pada konfigurasi 128 neuron:



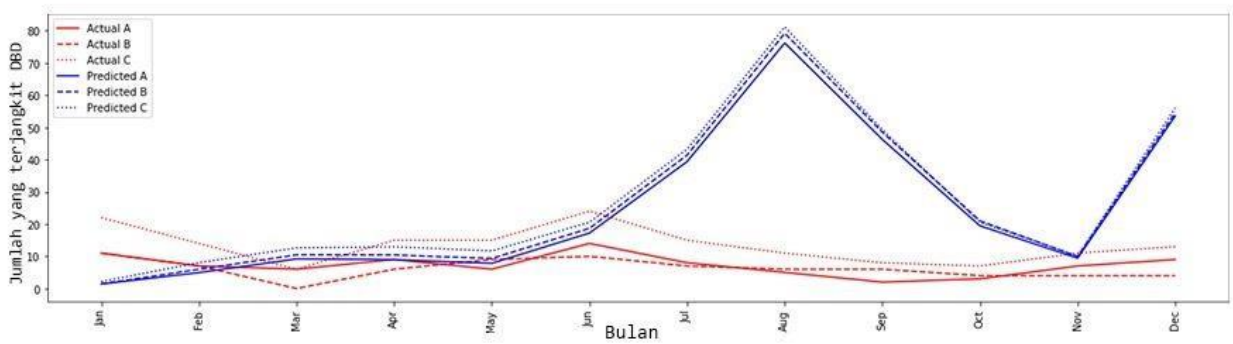
Gambar 4. 4 Grafik hasil percobaan 128 neuron dan dataset dengan preprocessing

Selanjutnya pada konfigurasi 64 neuron didapatkan hasil sebagai berikut:



Gambar 4. 5 Grafik hasil percobaan 64 neuron dan dataset dengan *preprocessing*

Terakhir pada konfigurasi 32 neuron didapatkan hasil sebagai berikut:



Gambar 4. 6 Grafik hasil percobaan 32 neuron dan dataset dengan *preprocessing*

Dari hasil percobaan ini didapatkan hasil error sebagai berikut:

Tabel 3. Hasil error untuk dataset dengan *preprocessing*

Konfigurasi	Error
128 neuron	5465.07
64 neuron	1422.98
32 neuron	2154.88

Dari hasil diatas, didapatkan korelasi sebagai berikut:

Tabel 4. Korelasi dari masing-masing daerah untuk dataset dengan *preprocessing*

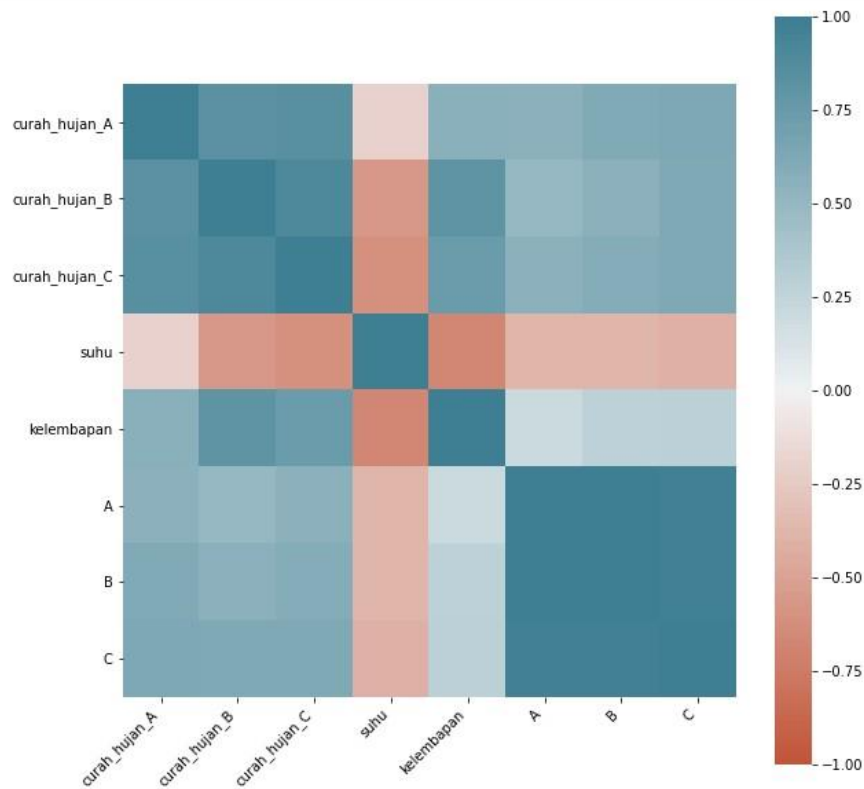
Daerah	Korelasi 128 neuron	Korelasi 64 neuron	Korelasi 32 neuron
Daerah A	0.472	0.238	-0.303
Daerah B	-0.220	-0.262	-0.159
Daerah C	0.319	0.015	-0.264

Untuk pengujian dataset dengan *preprocessing* didapatkan hasil bahwa jaringan LSTM dengan neuron 128 lebih baik daripada 64 dan 32 neuron.

4.3 Analisis Hasil Pengujian

Berdasarkan hasil eksperimen, ternyata hasil terbaik didapatkan pada konfigurasi 32 neuron dengan dataset tanpa *preprocessing*. Hal ini membuktikan bahwa LSTM dengan jumlah neuron yang lebih sedikit memiliki performa yang lebih baik daripada jumlah neuron yang lebih banyak, selain itu LSTM juga memiliki performa yang lebih baik pada dataset yang tidak di *preprocessing*. Hal sebaliknya terjadi pada dataset yang di *preprocessing*, hasilnya menunjukkan lebih baik yang menggunakan konfigurasi 128 neuron, hal ini membuktikan bahwa *preprocessing* dapat merusak informasi penting yang ada pada data, sehingga diperlukan jumlah neuron yang lebih banyak untuk menghasilkan prediksi yang lebih baik.

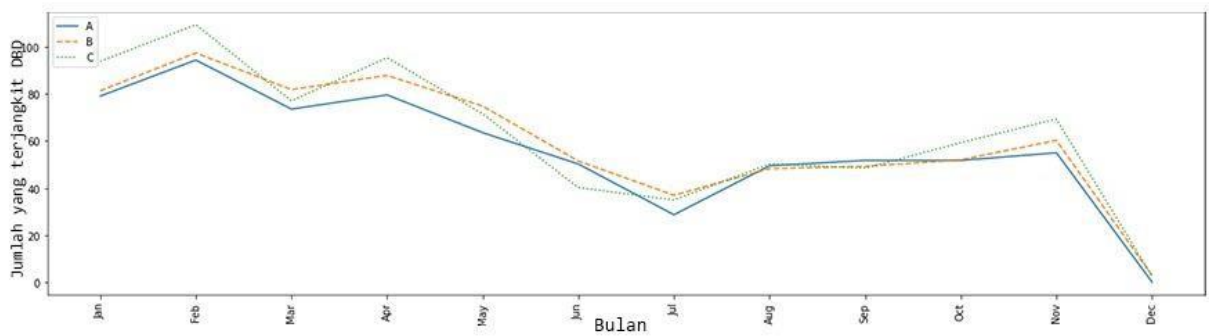
Disini penulis mencoba memvisualisasikan bagaimana hubungan antara data curah hujan, suhu, dan kelembaban dengan jumlah pasien yang terjangkit DBD tiap daerah, didapatkan hasil heatmap sebagai berikut:



Gambar 4.7 Hasil heatmap

Dari heatmap diatas, didapatkan kesimpulan ternyata data curah hujan memiliki korelasi yang positif dengan jumlah pasien yang terjangkit DBD, sedangkan data suhu memiliki korelasi negatif dan kelembaban memiliki korelasi yang mendekati nol. Data ini menandakan bahwa kenaikan curah hujan serta penurunan suhu dapat meningkatkan jumlah pasien DBD, sedangkan sebaliknya dapat mengurangi jumlah pasien DBD.

Sebagai perbandingan, penulis mencoba memvisualisasikan grafik rata-rata jumlah pasien yang terkena DBD tiap bulanan dari tahun 2015 hingga 2019, didapatkan hasil sebagai berikut:



Gambar 4. 8 Grafik rata-rata jumlah kasus DBD

Dari grafik diatas, di dapat informasi bahwa di awal tahun jumlah rata-rata pasien yang terkena DBD lebih tinggi daripada di akhir tahun, hal ini juga terjadi pada grafik hasil prediksi yang dihasilkan pada eksperimen dengan dataset yang sudah di *preprocessing*. Dari hasil grafik diatas didapatkan bahwa curah hujan juga mempengaruhi kenaikan jumlah pasien DBD, karena di Indonesia musim hujan terjadi pada bulan oktober hingga februari, dan terjadi kenaikan pada bulan oktober dan januari hingga februari.

5. Kesimpulan

Dalam tugas akhir ini, dilakukan percobaan dengan dataset yang tanpa di *preprocessing* dan di *preprocessing* terlebih dahulu dengan tiga model yang berbeda, yaitu model dengan 128 neuron, 64 neuron dan 32 neuron, didapatkan hasil bahwa dataset yang tanpa di *preprocessing* memiliki korelasi yang lebih baik daripada dataset yang tanpa di *preprocessing*. Selain itu jumlah neuron yang lebih sedikit pada dataset yang tanpa di *preprocessing* menghasilkan hasil yang lebih baik, sedangkan pada dataset yang di *preprocessing* memerlukan jumlah neuron yang lebih banyak untuk menghasilkan hasil yang terbaik. Dari hasil ini dapat disimpulkan bahwa kenaikan jumlah pasien di tiap daerah yang terjadi pada data validasi juga terjadi pada hasil prediksi begitu juga sebaliknya jika terjadi penurunan. Kenaikan jumlah pasien DBD cenderung terjadi pada bulan oktober hingga februari, hal ini disebabkan karena curah hujan yang meningkat pada masa tersebut yang mengakibatkan perkembangbiakan nyamuk *aedes aegypti* juga meningkat.

Dari hasil pengujian didapatkan konfigurasi terbaik pada dataset tanpa *preprocessing* dengan 32 neuron memiliki error 118.783 dan korelasi dari masing-masing daerahnya yaitu, Daerah A sebesar 0.594, Daerah B 0.420 serta Daerah C adalah sebesar 0.639. Jumlah kasus yang diprediksikan dari masing-masing daerah adalah sebagai berikut:

Tabel 5. Jumlah kasus DBD setaip bulan dari masing-masing daerah

Month / Daerah	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Daerah A	27	21	0	3	10	25	5	16	3	6	25	32
Daerah B	27	22	2	4	11	26	7	20	5	8	26	32
Daerah C	30	30	4	6	14	32	12	20	8	9	28	34

Saran dari penulis adalah dengan menggunakan data train yang lebih banyak serta fitur data yang lebih bervariasi agar bisa meningkatkan hasil prediksinya, selain itu penulis juga menyarankan menggunakan metode yang lebih baik yang mungkin saja ada di masa mendatang.

Referensi

- [1] R. YM, *Profil Dinas Kesehatan Kota Pagaram Tahun 2014*, no. 1. Pagaram: Departemen Kesehatan, 2014.
- [2] D. Nurainy, *Profil Kesehatan Sumatera Selatan Tahun 2015*. Provinsi Sumatera Selatan: Departemen Kesehatan, 2015.
- [3] Y. L. Hii, H. Zhu, N. Ng, L. C. Ng, and J. Rocklöv, "Forecast of Dengue Incidence Using Temperature and Rainfall," *PLoS Negl. Trop. Dis.*, vol. 6, no. 11, 2012.
- [4] R. N. Fadhlani, "Prediksi Pola Penyebaran Demam Berdarah Dengue di Kabupaten Bandung Menggunakan Algoritma Adaptive Neuro Fuzzy Inference System (ANFIS)," 2018.
- [5] G. Ramadan, "Prediksi Pola Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kabupaten Bandung Menggunakan Hybrid Genetic Algorithm dan Fuzzy," 2019.
- [6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [7] P. Sugiartawan, A. A. Jiwa Permana, and P. I. Prakoso, "Forecasting Kunjungan Wisatawan Dengan Long Short Term Memory (LSTM)," *J. Sist. Inf. dan Komput. Terap. Indones.*, vol. 1, no. 1, pp. 43–52, 2018.
- [8] N. S. Roslan, Z. A. Latif, and N. C. Dom, "Dengue Cases Distribution Based on Land Surface Temperature and Elevation," no. August, 2016.
- [9] I. B. Sudrajat, "PREDIKSI PENYEBARAN PENYAKIT DEMAM BERDARAH DENGAN MENGGUNAKAN ALGORITMA GENETIKA STUDI KASUS KOTA BANDUNG," 2017.
- [10] P. Manivannan, "Dengue Fever Prediction Using K-Means Clustering Algorithm," pp. 1–5, 2017.
- [11] M. M. Muzakki, "The Spreading Prediction of Dengue Hemorrhagic Fever (DHF) In Bandung Regency Using K-Means Clustering and Support Vector Machine Algorithm," 2018.
- [12] J. Soepardi, "Buletin Jendela Epidemiologi , Volume 2 , Agustus 2010," vol. 2, 2010.
- [13] K. Kesehatan, *InfoDatin-Situasi-Demam-Berdarah-Dengue*. Indonesia, 2017.
- [14] Q. Xiaoyun, K. Xiaoning, C. Zhang, S. Jiang, and M. Xiuda, "Short-term prediction of wind power based on deep Long Short-Term Memory," *Asia-Pacific Power Energy Eng. Conf. APPEEC*, vol. 2016-Decem, pp. 1148–1152, 2016.
- [15] M. Wildan Putra Aldi, Jondri, and A. Aditsania, "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin," *e-Proceeding Eng.*, vol. 5, no. 2, pp. 3548–3555, 2018.
- [16] Syilfi, D. Ispriyanti, and D. Safitri, "Analisis Regresi Linier Piecewise Dua Segmen," vol. 1, 2012.
- [17] A. S. Raksanagara, N. Arisanti, and F. Rinawan, "Dampak Perubahan Iklim Terhadap Kejadian Demam Berdarah Di Jawa-Barat," *J. Sist. Kesehat.*, vol. 1, no. 1, pp. 43–47, 2015.
- [18] I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. 2011.
- [19] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.