

Prediksi Penyebaran Kasus Demam Berdarah Berdasarkan Tingkat Keparahan di Provinsi DKI Jakarta Menggunakan Metode *Deep Learning*

Kusumastuti, Evelyn Anggita, Mauridhi Hery Purnomo, dan Eko Mulyanto Yuniarno
Departemen Teknik Komputer, Institut Teknologi Sepuluh Nopember (ITS)
e-mail : ekomulyanto@ee.its.ac.id

Abstrak—Demam berdarah *Dengue* atau DBD merupakan penyakit yang berasal dari infeksi salah satu dari empat virus *Dengue*, penularan terjadi melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Nyamuk ini sering ditemukan di daerah tropis dan subtropis. Indonesia, sebagai negara kesatuan yang terletak di Asia Tenggara merupakan negara yang beriklim tropis dimana penyakit DBD sudah menjadi penyakit endemik disetiap daerah, terutama pada daerah yang padat penduduk seperti di Jakarta menjadi habitat oleh berbagai jenis nyamuk *Aedes*. Untuk menekan terjadinya lonjakan kasus DBD maka diperlukan penanganan yang tepat baik dari masyarakat maupun pemerintah Provinsi DKI Jakarta. Oleh karena itu, tidak dapat dipungkiri pemanfaatan kemajuan teknologi juga dilibatkan seperti memprediksi daerah rawan lonjakan kasus DBD di Provinsi DKI Jakarta dengan menggunakan model *Deep Learning* ANN. Tujuan dari penggunaan metode ini adalah dapat menghasilkan prediksi yang lebih akurat terhadap daerah rawan peningkatan kasus demam berdarah *Dengue* guna mendapatkan tindakan preventif yang lebih. Hal ini dapat dilihat dari akurasi hasil prediksi probabilitas Stasiun Meteorologi Kemayoran sebesar 95,24; Stasiun Meteorologi Halim Perdana Kusuma sebesar 84,32; Stasiun Maritim Meteorologi Tanjung Priok sebesar 82,61.

Kata Kunci—Artificial Neural Network, Deep Learning, Demam Berdarah *Dengue*, Prediksi.

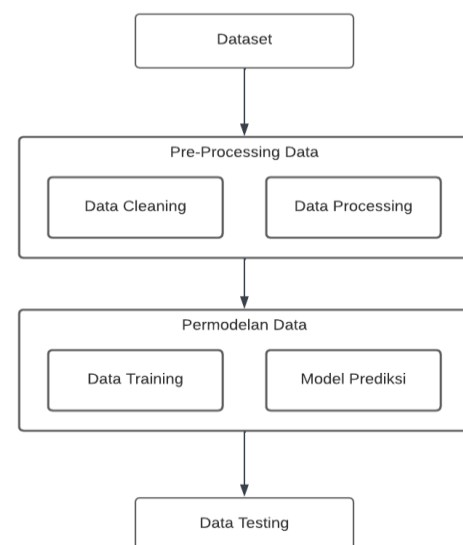
I. PENDAHULUAN

DEMAM berdarah *Dengue* atau DBD merupakan penyakit menular yang berasal dari infeksi virus *Dengue* melalui gigitan nyamuk betina *Aedes aegypti* dan *Aedes albopictus*. Virus dengue termasuk dalam genus *Flavivirus*, famili *Flaviviridae* dan mempunyai empat jenis serotipe, yaitu Den-1, Den-2, Den-3, Den-4. Penyakit ini kerap kali muncul pada wilayah beriklim tropis dan subtropis, seperti di Kawasan Asia Timur dan Asia Selatan. Selain kedua kawasan tersebut, kawasan Asia Tenggara yang juga memiliki iklim tropis, seperti Indonesia. Indonesia merupakan negara kesatuan yang dimana penyakit demam berdarah sudah menjadi penyakit endemik disetiap daerah. Tercatat bahwa setiap tahun tercatat laporan penambahan kasus demam berdarah. Pada tahun 2020, kasus demam berdarah mencapai 95.893 orang yang tersebar di 472 kota/kabupaten di 34 provinsi se-Indonesia.

Sedangkan kematian akibat penyakit demam berdarah mencapai 661 orang yang terjadi di 219 kota/kabupaten se-Indonesia. Penderita demam berdarah ini terdiri dari berbagai golongan umur, diantaranya penderita usia di bawah 1 tahun sebanyak 3,13%, 1-4 tahun sebanyak 14,88%, 5-14 tahun sebanyak 33,97%, 15-44 tahun sebanyak 37,45%, dan usia di atas 44 tahun sebanyak 11,57%. Sedangkan, proporsi kematian akibat demam berdarah pada umur di bawah 1 tahun

Tabel 1.
Spesifikasi *Personal Computer* yang digunakan

Processor	11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz 2.80 GHz
Memory	16.00 GB
Storage	SSD 1,9 TB
Graphics Card	Intel® Iris® Xe Graphics
Operating System	Windows 11 Home Single Language



Gambar 1. Blok Diagram Desain Sistem.

sebanyak 10,32%, 1-4 tahun sebanyak 28,57%, 5-14 tahun sebanyak 34,13%, 15-44 tahun sebanyak 15,87%, dan kematian di atas usia 44 tahun sebanyak 11,11%. Terutama di Provinsi DKI Jakarta selama tahun 2020 terdapat 970 kasus demam berdarah.

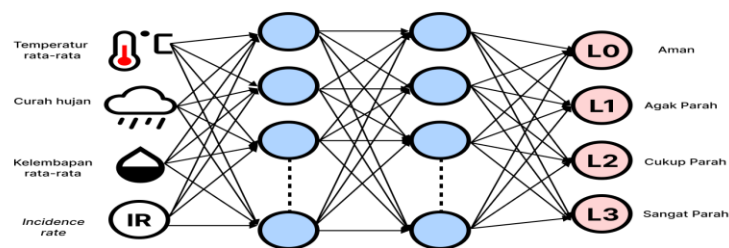
Melihat dari kondisi tersebut. Pemerintah Provinsi DKI Jakarta melakukan berbagai upaya untuk meningkatkan kesadaran masyarakat terhadap bahayanya penyakit demam berdarah ini melalui berbagai penyuluhan. Selain itu, pemerintah juga telah membentuk juru pemantau jumantik pada tiap RT guna melakukan upaya pencegahan terbentuknya habitat nyamuk di daerah tersebut. Upaya lain juga dilakukan dengan menerapkan Pemberantasan Sarang Nyamuk (PSN) 3M Plus. M yang pertama adalah dengan menguras tempat yang sering digunakan sebagai penampungan air, seperti bak mandi dan toren air. M yang kedua adalah menutup rapat tempat-tempat penampungan air dan mengubur barang bekas yang berpotensi menjadi habitat baru nyamuk *Aedes aegypti* di dalam tanah. M yang ketiga adalah dengan memanfaatkan kembali limbah barnag bekas yang memiliki nilai ekonomis atau daur ulang dan pada poin

Tabel 2.
Jumlah Kasus Demam Berdarah di Provinsi DKI Jakarta Tahun 2016-2020

	Jakarta Barat	Jakarta Pusat	Jakarta Selatan	Jakarta Timur	Jakarta Utara	Kep. Seribu
2016	880	718	970	981	639	140
2017	822	129	627	1.265	518	1
2018	909	115	673	846	271	8
2019	2.305	492	1.976	3.004	271	6
2020	3.416	326	830	1.361	542	6
Rata-Rata	1.666	356	1.015	1.491	448	32

Tabel 3.
Uji Korelasi antara Data Iklim dengan Jumlah Penderita DBD

	Jakarta Barat	Jakarta Pusat	Jakarta Selatan	Jakarta Timur	Jakarta Utara	Kep. Seribu
Tavg	0,0249	0,0252	-0,0896	-0,0119	-0,0044	0,0665
RHavg	0,1347	0,3976	-0,0194	-0,0653	0,2511	0,1933
RR	-0,0872	-0,0413	-0,1454	-0,0943	-0,0717	0,1400



Gambar 2. Arsitektur ANN pada permodelan data kasus demam berdarah.

Plus adalah dengan melakukan bentuk upaya pencegahan tambahan, seperti dengan memelihara ikan pemakan jentik nyamuk, menggunakan obat anti nyamuk, dan membersihkan lingkungan sekitar. Namun, upaya tersebut masih dianggap kurang, sehingga dibutuhkan cara lain untuk meningkatkan tingkat keberhasilan dalam mencegah kasus luar biasa demam berdarah.

Melihat kondisi tersebut, Pemprov DKI Jakarta telah melakukan berbagai upaya pencegahan dengan memanfaatkan teknologi, seperti model prediksi kasus, sebagai alat pencegahan untuk daerah rawan lonjakan kasus demam berdarah. Namun, model prediksi yang dimiliki pemerintah hanya menggunakan satu parameter. Sehingga diperlukan lebih dari satu parameter supaya dapat dilakukan pencegahan secara maksimal. Parameter yang digunakan untuk melengkapi parameter yang ada adalah dengan menambahkan parameter curah hujan, suhu, dan tingkat kejadian. Kemudian memperluas cakupan prediksi untuk seluruh Provinsi Jakarta, yaitu Kabupaten Kepulauan Seribu.

Metode *Deep Learning* merupakan gabungan ilmu dari *Artificial Intelligence* (AI) dan *Machine Learning* yang dapat mengabstraksi data tingkat tinggi menggunakan fungsi transformasi non-linear. Model *Deep Learning* adalah pengembangan dari jaringan saraf tiruan berlapis atau *Neural Network Multiple Layer*. Salah satu cabang dari *Deep Learning* adalah *Artificial Neural Network* (ANN) dengan algoritma *Backpropagation*.

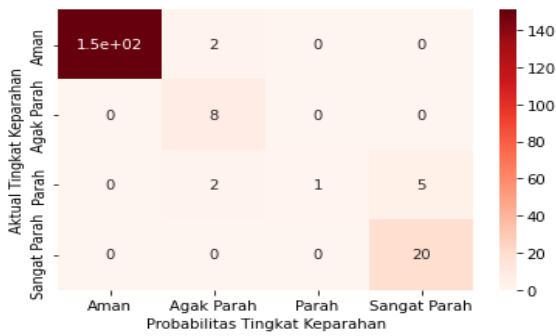
A. Artificial Neural Network

Artificial Neural Network atau jaringan saraf tiruan yang meniru cara kerja *Neuron* atau saraf otak manusia yang dapat memberikan simulasi berupa rangsangan, melakukan proses, toleransi terhadap *error*, *parallel processing*, dan

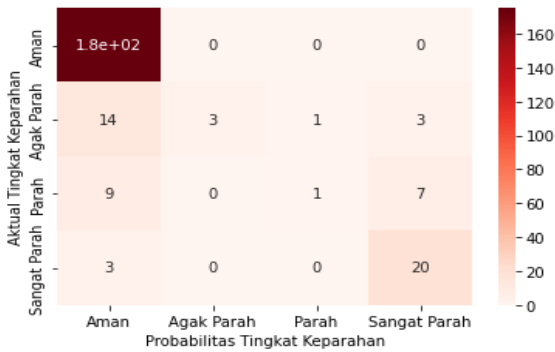
memberikan tanggapan. Arsitektur dari ANN terdiri dari *input*, *hidden layers*, dan *output*. Di dalam struktur ANN terdapat *two-layer network* yang dapat disebut dengan *perceptron*. Fungsi *perceptron* untuk melakukan pekerjaan klasifikasi dalam pembelajaran tertentu dengan menambah bobot pada setiap koneksi antar *network*. Metode ini sering digunakan untuk pengklasifikasian pola yang didapatkan dari *input* ke pola baru pada *output*, penyimpanan pola akan dipanggil kembali, pemetaan pola-pola yang sejenis, pengoptimasi permasalahan, dan prediksi. *Artificial Neural Network* memiliki beberapa struktur model, salah satunya adalah *Feedforward Neural Network*. *Neural Network* dengan struktur *Feedforward* mempunyai karakteristik pada proses data diolahnya yaitu proses ini tidak terjadi pengulangan pembelajaran (*loop*) dari *layer output* menuju *layer input*. *Neural Network* yang menggunakan struktur *feedforward*, diantaranya adalah *single-layer perceptron*, *multi-layer perceptron*, *radial-basis function networks*, dan *backpropagation neural network*.

B. Backpropagation

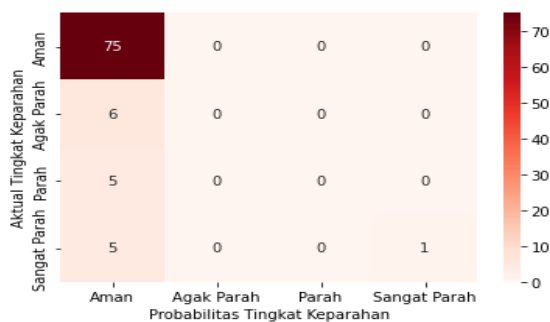
Backpropagation adalah algoritma pembelajaran dalam *neural network* yang berfungsi untuk memperkecil tingkat *error* dengan cara menyesuaikan bobotnya yang didasarkan pada perbedaan output dan target yang diinginkan. Algoritma ini merupakan sebuah metode sistematis yang digunakan sebagai pelatihan *multi-layer neural network* dan merupakan perkembangan dari *single-layer neural network*. *Hidden layers* yang terdapat pada arsitektur *backpropagation* memiliki tingkat *error* yang lebih kecil dibandingkan dengan tingkat *error* pada *single layer neural network*. Hal ini berfungsi sebagai wadah untuk melakukan perbaruan, dan



Gambar 3. Pengujian *Confusion Matrix* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Kemayoran.



Gambar 4. Pengujian *Confusion Matrix* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Halim Perdana Kusuma.



Gambar 5. Pengujian *Confusion Matrix* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Maritim Tanjung Priok.

menyesuaikan bobot sehingga bobot baru yang didapatkan bisa diarahkan mendekati target output yang diinginkan.

II. PENELITIAN TERKAIT

Penelitian tentang prediksi kasus demam berdarah pernah dilakukan sebelumnya, seperti penelitian yang dilakukan oleh Ichwani, et.all (2019) tentang pembuatan model prediksi menggunakan metode *Extreme Machine Learning* (ELM) dengan parameter dataset iklim dan dataset kasus demam berdarah [1]. Selain itu, sasaran ruang lingkup penelitian berada di kota Semarang. Hasil yang didapatkan dari penelitian tersebut adalah pemilihan nilai kinerja model ELM terbaik sebesar 0,0116 dengan n split 13, jumlah *hidden neuron* sebanyak 6, dan waktu training yang kurang dari satu detik.

Penelitian yang serupa juga pernah dilakukan oleh (Respati, Wanti, & Nindrea, 2020) mengenai prediksi kasus demam berdarah *dengue* di Kupang [2]. Penelitian tersebut menggunakan parameter data yang sama, yaitu temperatur rata-rata, kelembapan rata-rata, curah hujan, serta *incidence*

Tabel 4. Pengujian *Classification Report* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Kemayoran.

	Precision	Recall	F1-Score	Support
L0_prob	1.00	0.99	0.99	153
L1_prob	0.67	1.00	0.80	8
L2_prob	1.00	0.12	0.22	8
L3_prob	0.80	1.00	0.89	20
Accuracy			0.95	189
Macro avg	0.87	0.78	0.73	189
Weighted avg	0.96	0.95	0.94	189

Tabel 5. Pengujian *Classification Report* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Halim Perdana Kusuma

	Precision	Recall	F1-Score	Support
L0_prob	0.87	1.00	0.93	175
L1_prob	1.00	0.14	0.25	21
L2_prob	0.50	0.06	0.11	17
L3_prob	0.67	0.87	0.75	23
Accuracy			0.84	236
Macro avg	0.76	0.52	0.51	236
Weighted avg	0.84	0.84	0.79	236

Tabel 6. Pengujian *Classification Report* terhadap Data Prediksi Probabilitas Stasiun Meteorologi Maritim Tanjung Priok.

	Precision	Recall	F1-Score	Support
L0_prob	0.82	1.00	0.90	75
L1_prob	0.00	0.00	0.00	6
L2_prob	0.00	0.00	0.00	5
L3_prob	1.00	0.17	0.29	6
Accuracy			0.83	92
Macro avg	0.46	0.29	0.30	92
Weighted avg	0.74	0.83	0.76	92

rate demam berdarah yang terjadi di Kupang pada tahun 2010-2019. Penelitian ini menggunakan metode *autoregressive integrated moving average* (ARIMA). Prediksi yang dibuat berupa prediksi jumlah kasus demam berdarah pada tahun 2020 dimana didapatkan hasil bahwa prediksi yang dihasilkan mengalami penurunan satu kasus pada setiap kuartar tahun. Namun, prediksi ini masih belum dapat menampilkan tingkat keparahan dari jumlah kasus demam berdarah yang ada.

III. METODOLOGI

Pada pengerjaan penelitian ini dilakukan menggunakan *Personal Computer* atau PC dengan spesifikasi seperti pada Tabel 1.

A. Desain Sistem

Desain sistem merupakan penggambaran, perencanaan, dan pembuatan sketsa atau pengaturan dari berbagai elemen yang berbeda ke dalam satu kesatuan yang utuh dan memiliki fungsi. Desain sistem ini sangat penting karena menjadi gambaran besar terhadap tahapan yang dilakukan serta dapat mempermudah dalam pelaksanaan penelitian ini. Desain sistem yang digunakan tertera pada Gambar 1. Gambar 1 merupakan alur teknis dari desain sistem penelitian ini.

Tahapan dalam penelitian ini adalah mengambil dataset, kemudian dilakukan *pre-processing* data yang berisi data *cleaning* dan data *processing*. Data *cleaning* dilakukan pada dataset mentah, yaitu dengan mengisi *missing value* pada setiap dataset yang ada. Kemudian dilakukan pemrosesan

Tabel 7.
Contoh Data Prediksi Probabilitas dan Pelabelan Stasiun Meteorologi Kemayoran.

Bulan Tahun	Kota Administrasi	Kecamatan	L0_prob	L1_prob	L2_prob	L3_prob	Label
2020-04-30	Jakarta Barat	Kebon Jeruk	0,52	0,40	0,07	0,01	AMAN
2020-04-30	Jakarta Barat	Kembangan	0,00	0,00	0,00	1,00	SANGAT PARAH
2020-04-30	Jakarta Barat	Palmerah	0,00	0,04	0,12	0,84	SANGAT PARAH
2020-04-30	Jakarta Barat	Taman Sari	0,00	0,00	0,00	1,00	SANGAT PARAH
2020-04-30	Jakarta Barat	Tambora	0,21	0,58	0,16	0,05	AGAK PARAH
2020-05-31	Jakarta Pusat	Cempaka Putih	0,01	0,37	0,38	0,24	PARAH
2020-05-31	Jakarta Pusat	Gambir	0,87	0,13	0,00	0,00	AMAN
2020-05-31	Jakarta Pusat	Johar Baru	0,77	0,22	0,01	0,00	AMAN
2020-05-31	Jakarta Pusat	Kemayoran	0,00	0,01	0,12	0,87	SANGAT PARAH
2020-05-31	Jakarta Pusat	Menteng	0,99	0,01	0,00	0,00	AMAN

dengan menyatukan ketiga dataset menjadi satu dataset yang siap untuk diproses. Pada tahap data *processing* dilakukan pembuatan parameter baru yang akan digunakan sebagai penentu tingkat keparahan klasifikasi. Selanjutnya pada tahap permodelan data yang dibagi menjadi dua tahap, yaitu data *training* dan membuat model prediksi. Tahap akhir dari penelitian ini adalah pengujian data *testing*, data prediksi probabilitas, dan pelabelan data berdasarkan parameter tingkat keparahan.

Dataset yang digunakan dalam pembuatan model prediksi ini, yaitu dataset iklim yang terdiri dari tiga faktor: suhu rata-rata, curah hujan, dan kelembaban rata-rata periode waktu 2016-2020 berdasarkan stasiun meteorologi yang mencakup seluruh provinsi DKI Jakarta. Sehingga daerah yang berada pada stasiun meteorologi yang sama, maka akan memiliki data iklim yang sama, misalnya, data iklim di Jakarta Pusat dan Jakarta Barat yang berada di wilayah cakupan Stasiun Meteorologi Kemayoran, memiliki suhu rata-rata 28,6°C, kelembaban rata-rata 75,9%, dan curah hujan 518,2 mm. Jakarta Timur dan Jakarta Selatan berada pada cakupan Stasiun Meteorologi Halim Perdana Kusuma, memiliki suhu rata-rata 28,2°C, kelembaban rata-rata 75,6%, dan curah hujan 950,6 mm. Jakarta Utara dan Kepulauan Seribu berada pada cakupan Stasiun Meteorologi Maritim Tanjung Priok memiliki suhu rata-rata 28,8°C, kelembaban rata-rata 76%, dan curah hujan 531,2 mm.

B. Dataset

Penelitian ini menggunakan dataset iklim, jumlah kasus demam berdarah, dan jumlah populasi di Provinsi DKI Jakarta. Dataset iklim yang digunakan terdiri dari tiga tipe data, yaitu suhu rata-rata (TAvg) dalam satuan *celcius* (°C), curah hujan (RR) dalam satuan milimeter (mm), dan kelembaban rata-rata (RH_avg) dalam satuan persen (%). Ketiga data tersebut diambil di setiap stasiun meteorologi yang terletak di provinsi DKI Jakarta. Hal ini dikarenakan Provinsi DKI Jakarta memiliki kondisi geografis yang berbeda, dimana pada Kota Jakarta Utara merupakan daerah pesisir, Kabupaten Kepulauan Seribu merupakan daerah kepulauan, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Timur merupakan daerah dataran rendah, dan Kota

Jakarta Selatan merupakan daerah perbukitan. Data kedua yang digunakan adalah dataset jumlah kasus demam berdarah di DKI Jakarta dalam kurun waktu 2016-2020 serta dataset jumlah penduduk berdasarkan usia dan jenis kelamin perkelurahan dalam kurun waktu 2016-2020 yang diambil melalui *Opendata* Jakarta. Data jumlah kasus demam berdarah di Provinsi DKI Jakarta dapat dilihat pada Tabel 2. Berdasarkan Tabel 2 dapat diketahui bahwa :

1. Kota Jakarta Barat memiliki rata-rata jumlah penderita pertahun mencapai 1666 orang.
2. Kota Jakarta Pusat memiliki rata-rata jumlah penderita pertahun sebanyak 356 orang.
3. Kota Jakarta Selatan memiliki rata-rata jumlah penderita pertahun sebanyak 1015 orang.
4. Kota Jakarta Timur memiliki rata-rata jumlah penderita pertahun sebanyak 1491 orang.
5. Kota Jakarta Utara memiliki rata-rata jumlah penderita pertahun sebanyak 448 orang.
6. Kabupaten Kepulauan Seribu memiliki rata-rata jumlah penderita pertahun sebanyak 32 orang.

Sebelum dilakukan *pre-processing* data, dilakukan pengujian terhadap korelasi antara dataset iklim (suhu, curah hujan, dan kelembaban) dengan jumlah kasus demam berdarah menggunakan metode *bivariate pearson*, seperti pada Tabel 3.

Analisis *bivariate pearson* yang dilakukan mendefinisikan bahwa data iklim Kota Jakarta Barat memiliki korelasi dengan jumlah penderita DBD, sebagai berikut :

1. Dataset suhu memiliki korelasi sangat rendah, arah positif.
2. Dataset kelembapan udara memiliki korelasi sangat rendah, arah positif.
3. Dataset curah hujan memiliki korelasi sangat rendah, arah negatif.

Sehingga dapat diartikan ketika suhu dan kelembapan udara naik, sedangkan intensitas curah hujan turun. Maka, dapat terjadi kenaikan jumlah penderita DBD, namun sangat rendah. Kota Jakarta Pusat dengan korelasi sebagai berikut : (1) Dataset suhu memiliki korelasi sangat rendah, arah positif; (2) Dataset kelembapan udara memiliki korelasi rendah, arah positif; (3) Dataset curah hujan memiliki

korelasi sangat rendah, arah negatif.

Sehingga dapat diartikan ketika suhu dan kelembapan udara naik, sedangkan intensitas curah hujan turun. Maka, dapat terjadi kenaikan jumlah penderita DBD, dengan probabilitas yang rendah. Kota Jakarta Selatan dengan korelasi sebagai berikut :

1. Dataset suhu memiliki korelasi sangat rendah, arah negatif.
2. Dataset kelembapan udara memiliki korelasi sangat rendah, arah negatif.
3. Dataset curah hujan memiliki korelasi sangat rendah, arah negatif.

Sehingga dapat diartikan ketika suhu, kelembapan udara, dan intensitas curah hujan turun. Maka, dapat terjadi kenaikan jumlah penderita DBD, namun sangat rendah.

Kota Jakarta Timur dengan korelasi sebagai berikut :

1. Dataset suhu memiliki korelasi sangat rendah, arah negatif.
2. Dataset kelembapan udara memiliki korelasi sangat rendah, arah negatif.
3. Dataset curah hujan memiliki korelasi sangat rendah, arah negatif.

Sehingga dapat diartikan ketika suhu, kelembapan udara, dan intensitas curah hujan turun. Maka, dapat terjadi kenaikan jumlah penderita DBD, namun sangat rendah.

Kota Jakarta Utara dengan korelasi sebagai berikut :

1. Dataset suhu memiliki korelasi sangat rendah, arah negatif.
2. Dataset kelembapan udara memiliki korelasi rendah, arah positif.
3. Dataset curah hujan memiliki korelasi sangat rendah, arah negatif.

Sehingga dapat diartikan ketika suhu dan intensitas curah hujan turun, sedangkan kelembapan udara naik. Maka, dapat terjadi kenaikan jumlah penderita DBD, namun rendah. Kabupaten Kepulauan Seribu dengan korelasi sebagai berikut :

1. Dataset suhu memiliki korelasi sangat rendah, arah positif.
2. Dataset kelembapan udara memiliki korelasi sangat rendah, arah positif.
3. Dataset curah hujan memiliki korelasi sangat rendah, arah positif.

Sehingga dapat diartikan ketika suhu, kelembapan udara, dan intensitas curah hujan naik. Maka, dapat terjadi kenaikan jumlah penderita DBD, namun sangat rendah.

C. Data pre-processing

Pada tahap ini dilakukan pengolahan terhadap dataset yang telah disatukan berdasarkan lingkup stasiun meteorologi yang menaunginya. Untuk menghindari terjadinya *error* atau *overlap* data maka pengolahan dilakukan secara terpisah juga berdasarkan stasiun yang menaunginya. Dalam membuat model prediksi, maka diperlukan parameter untuk mengklasifikasi tingkat keparahan akan kasus demam berdarah. Oleh karena itu, dibuat parameter *Incidence Rate*. Pada perhitungan *Incidence Rate* menggunakan konstanta "k" diberi nilai 10.000 yang memiliki arti perhitungan *Incidence Rate* ini dilakukan per 10.000 penduduk. Sehingga didapatkan persamaan (1).

$$\text{Incidence Rate} = \frac{\text{dataset penderita dbd}}{\text{jumlah penduduk}} \times 10.000 \quad (1)$$

Perhitungan diatas digunakan untuk mengklasifikasikan hasil prediksi menjadi empat tingkat keparahan.

D. Permodelan Data

Pada tahap permodelan data menggunakan dataset siap proses, dimana dataset mentah telah melalui proses *pre-processing* seperti pada jaringan ANN yang digunakan pada penelitian yang dapat dilihat pada Gambar 2.

Penelitian ini menggunakan empat input, yaitu suhu rata-rata, curah hujan rata-rata, kelembapan rata-rata, dan *Incidence Rate* yang kemudian akan diolah di dalam 2 *hidden layers* dan akan menghasilkan empat klasifikasi produk luaran berdasarkan level keparahan, sebagai berikut :

1. L0 atau berlabel aman dengan kondisi nilai prediksi yang dihasilkan dibawah 0,5.
2. L1 atau berlabel agak parah dengan kondisi nilai prediksi yang dihasilkan diantara rentang 0,5-0,75.
3. L2 atau berlabel parah dengan kondisi nilai prediksi yang dihasilkan diantara rentang 0,75-1,00.
4. L3 atau berlabel sangat parah dengan kondisi nilai prediksi yang dihasilkan di atas 1,00.

Pada tahap permodelan data dilakukan dua langkah, yaitu data *training* dan pembuatan model prediksi. Pada tahap data *training* ini, hal pertama yang dilakukan adalah dengan mengubah tipe data pada kolom "Bulan Tahun" yang masih berbentuk *object* menjadi tipe data *date_time*. Hal ini dikarenakan dataset pada kolom Bulan_Tahun digunakan sebagai acuan pengolahan dataset. Kemudian dilakukan pembagian data menjadi data *training* dan data *testing*, yaitu sebesar 80% dari data keseluruhan dan 20% data tersebut akan digunakan untuk data *testing*. Data *training* dibuat lebih besar, supaya model *deep learning* dapat lebih banyak mempelajari data yang diberikan, sehingga dapat menghasilkan prediksi yang lebih baik. Setelah dilakukan pembagian data *training* dan data *testing*, diperlukan pendeklarasian parameter yang akan digunakan sebagai data *input*. Parameter dataset tersebut adalah dataset Tavg, RH avg, RR, dan IR yang didefinisikan sebagai nilai X_train. Selain pendeklarasian parameter dataset input, perlu juga dilakukan pendeklarasian parameter dataset *output* yang terdiri dari L0, L1, L2, dan L3 sebagai nilai Y_train. Parameter *input* dan *output* ini dideklarasikan pada setiap stasiun meteorologi yang digunakan.

Pada tahap model prediksi ini menggunakan *deep learning* dengan metode *artificial neural network backpropagation*. Dalam proses permodelannya menggunakan *library keras* dengan model *sequential*. Model *sequential* digunakan untuk struktur *layer* biasa dimana pada prosesnya memiliki satu tensor input dan satu tensor *output*. Model prediksi ini menggunakan empat parameter yang telah dideklarasikan sebelumnya, sebagai X_train. Keempat parameter ini, kemudian akan dilakukan *training* pada dua *hidden layers*. Dimana pada setiap *hidden layers* terdapat 1000 neuron dengan *activation functions* ReLU. *Activation function* ReLU memiliki komputasi yang sederhana dan mudah digunakan untuk memproses nilai input, dimana ketika nilai data *input* bernilai negatif, maka model akan mengartikan nilai data tersebut sebagai 0, sedangkan ketika nilai data *input* bernilai positif, maka model akan mengartikan sebagai 1. Pada *output*

layer menghasilkan empat nilai *output*, dimana *output* tersebut berdasarkan parameter *Y*, serta *activation functions* yang digunakan adalah *softmax*. Pada *output layer* ini menggunakan *activation functions* yang berbeda dengan *activation functions* yang berada pada *hidden layers*.

Hal ini dikarenakan dalam *softmax* akan mengubah vektor bilangan yang telah di-*input*-kan sebelumnya, menjadi vektor peluang, dimana peluang yang dihasilkan pada setiap nilai sebanding dengan skala relatif dari setiap nilai di dalam vektor tersebut. Untuk menghasilkan model yang telah didefinisikan maka model tersebut perlu di *compile* dengan beberapa syarat, diantaranya dengan menggunakan salah satu *loss functions*, yaitu *categorical crossentropy*. Tipe *loss functions* dipilih karena pada penelitian ini digunakan untuk melakukan tugas klasifikasi *multi-class*. *Categorical crossentropy* dirancang untuk mengukur perbedaan antara dua distribusi probabilitas. Kemudian untuk *optimizer* menggunakan algoritma Adam. Adam atau *Adaptive Moment Estimation* merupakan *adaptive learning method* yang dapat menghitung *learning rate* individu pada parameter yang berbeda karena algoritma ini menggunakan estimasi gradien pertama dan kedua untuk mengadaptasi *learning rate* pada setiap bobot *neural network* dan juga menggunakan matriks *accuracy* sebagai model evaluasi. Untuk mencegah *overfitting* model prediksi, maka digunakan *callbacks.EarlyStopping* yang berfungsi untuk memantau nilai *loss*. Ketika model dijalankan didapatkan nilai *loss* paling rendah dan tidak terjadi kenaikan nilai *loss* pada 5 *epoch* sesudahnya, maka model tersebut akan berhenti. *Epoch* yang digunakan pada penelitian ini sebanyak 100 *epochs* dan 1 *batch sizes*. *Epochs* disini dimaksudkan bahwa dataset pada *neural network* telah melalui satu putaran selama proses training, sehingga dapat diartikan bahwa model yang dibuat akan mempelajari data tersebut sebanyak 100 putaran. Dikarenakan 100 *epochs* terlalu besar untuk diolah komputer, maka perlu dikerucutkan lagi menjadi satuan-satuan kecil (*Batch*). Model ini menggunakan 1 *batch sizes* yang berarti dalam proses training menggunakan 1 sampel data yang disebar dalam *neural network* guna memudahkan *training* data.

Setelah dilakukan pembuatan model, kemudian dilakukan evaluasi model terhadap data *training* yang telah ditentukan sebelumnya menggunakan *accuracy*. Jika *accuracy* yang dihasilkan dianggap sudah baik, maka model prediksi ini dapat disimpan untuk dilakukan pengujian pada tahap selanjutnya. Pada penelitian ini menggunakan model prediksi yang sama yang diterapkan pada seluruh data berdasarkan stasiun meteorologi yang menaunginya.

E. Data Testing

Pada tahapan pengujian ini memerlukan dataset keseluruhan dan model prediksi sesuai dengan stasiun meteorologi yang telah dibuat sebelumnya. Pengujian model ini menggunakan dataset yang sama. Hal ini berkaitan dengan keterbatasan data yang dipublikasikan oleh *website* resmi *data.jakarta.go.id*. Karena dataset yang digunakan masih menjadi satu kesatuan, maka perlu dilakukan pembagian antara data training dengan data *testing* dengan perbandingan 80% digunakan sebagai nilai data *training* dan 20% sebagai data *testing*. Namun, pada tahap ini difokuskan terhadap data *testing*. Dalam melakukan *testing* dibutuhkan mendefinisikan

parameter *input* yang akan digunakan, yaitu dataset Tavg, RH_avg, RR, dan IR didefinisikan sebagai *X_test* dan parameter *output* yang akan dihasilkan, yaitu L0, L1, L2, dan L3 yang didefinisikan sebagai *Y_test*. Setelah dataset di *testing* menggunakan model prediksi, maka dilakukan pengujian performa model *testing* terhadap data aktual *X_test* dan *Y_test*. Langkah selanjutnya adalah pembuatan prediksi berdasarkan model yang sudah di lakukan *testing*. Prediksi yang dibuat menggunakan parameter *X_test* sebagai acuan *input*. Hasil prediksi yang didapatkan parameter kondisi tingkat keparahan kasus demam berdarah dalam format numerik. Untuk memudahkan pembacaan hasil prediksi kondisi tingkat keparahan kasus demam berdarah, maka hasil prediksi perlu diubah menjadi data *string*. Sehingga dibutuhkan suatu sintaks untuk memberikan label pada hasil prediksi dan menyimpan hasil prediksi yang didapatkan. Tidak hanya data *training* yang dilakukan pengujian performa, pengujian ini juga dilakukan terhadap hasil prediksi. Untuk pengujian performa hasil prediksi menggunakan *evaluation matrix* berupa *confusion matrix* dikarenakan model yang digunakan merupakan *multi-class classification*.

IV. HASIL DAN PENGUJIAN

Pada penelitian ini, dijelaskan mengenai hasil dari analisis dan pengujian yang dilakukan. Data yang digunakan untuk pengujian menggunakan dataset kombinasi yang didapatkan dari dataset iklim, dataset jumlah kasus demam berdarah, dan dataset jumlah penduduk di Provinsi DKI Jakarta. Untuk dataset iklim diambil dari beberapa stasiun meteorologi yang berada di Provinsi DKI Jakarta, yaitu : (1)Stasiun Meteorologi Kemayoran yang mencakup data iklim Kota Jakarta Pusat dan Jakarta Barat;(2)Stasiun Meteorologi Halim Perdana Kusuma yang mencakup data iklim Kota Jakarta Timur dan Jakarta Selatan;(3)Stasiun Meteorologi Maritim Tanjung Priok yang mencakup data iklim Kota Jakarta Utara dan Kabupaten Kepulauan Seribu.

A. Potensi Penyakit Demam Berdarah di Provinsi DKI Jakarta

Suhu rata-rata di Kota Jakarta Pusat dan Kota Jakarta Barat pada tahun 2016-2020 sebesar 28,6°C. Kedua kota ini memiliki suhu rata-rata yang sama karena data iklim yang digunakan berasal dari stasiun meteorologi yang sama, yaitu Stasiun Meteorologi Kemayoran. Hal ini juga berlaku pada stasiun meteorologi lainnya, pada Stasiun Meteorologi Halim Perdana Kusuma mencakup Kota Jakarta Timur dan Kota Jakarta Selatan pada tahun 2016-2020 dengan suhu rata-rata sebesar 28,2 °C, dan Stasiun Meteorologi Maritim Tanjung Priok mencakup Kota Jakarta Utara dan Kabupaten Kepulauan Seribu pada tahun 2016-2020 dengan suhu rata-rata sebesar 28,8 °C. Suhu ini merupakan suhu optimal untuk nyamuk *Aedes aegypti* berkembang biak.

Curah hujan rata-rata di Kota Jakarta Pusat dan Kota Jakarta Barat pada tahun 2016-2020 sebesar 518,2 mm. Sedangkan Kota Jakarta Timur dan Kota Jakarta Selatan selama tahun 2016-2020 memiliki intensitas curah hujan sebesar 950,6. Kota Jakarta Utara dan Kabupaten Kepulauan Seribu pada tahun 2016-2020 memiliki curah hujan sebesar 531,2 mm. Ketika kondisi curah hujan ini termasuk dalam

kategori curah hujan dengan intensitas tinggi (500 mm). Intensitas curah hujan yang tinggi ini dapat memicu berkembang biak nyamuk *Aedes aegypti*.

Kelembapan udara rata-rata di Kota Jakarta Pusat dan Kota Jakarta Barat pada periode waktu 2016-2020 memiliki kelembapan udara rata-rata sebesar 75,9%. Sedangkan kelembapan udara rata-rata di Kota Jakarta Timur dan Kota Jakarta Selatan sebesar 75,6%, dan kelembapan udara rata-rata di Kota Jakarta Utara dan Kabupaten Kepulauan Seribu mencapai 76%. Ketiga kondisi merupakan kelembapan udara yang optimal bagi nyamuk untuk berkembang biak.

B. Pengujian Akurasi

Pengujian akurasi ini dilakukan terhadap data *training* dan hasil model prediksi untuk mengetahui tingkat keakuratan model prediksi yang dibuat.

C. Model Prediksi Stasiun Meteorologi Kemayoran

Pada pemrosesan model prediksi, diperlukan pembagian dataset Stasiun Meteorologi Kemayoran yang memiliki 944 jumlah data, menjadi 80% untuk data *training* sebesar 755 data dan 20% untuk data *testing* sebesar 189 data. Data *training* kemudian dilakukan permodelan dengan 100 *epochs* didapatkan akurasi sebesar 93,77 dengan *loss* 19,20. Berdasarkan hasil akurasi dan *loss* yang didapatkan menunjukkan bahwa model prediksi pada *training* sudah cukup baik dikarenakan nilai akurasi yang tinggi dan mendekati 100.

D. Model Prediksi Stasiun Meteorologi Halim Perdana Kusuma

Pembuatan model prediksi untuk dataset Stasiun Halim Perdana Kusuma, diperlukan pembagian dataset Stasiun Halim Perdana Kusuma yang memiliki 1180 data, menjadi 80% untuk data *training* sebesar 944 data dan 20% untuk data *testing* sebesar 236 data. Data *training* ini digunakan untuk permodelan dengan 100 *epochs* dan didapatkan akurasi sebesar 89,83 dengan *loss* 27,55. Dari hasil nilai akurasi dan *loss* menunjukkan bahwa model prediksi pada *training* sudah cukup baik dikarenakan nilai akurasi yang tinggi dan mendekati 100.

E. Model Prediksi Stasiun Meteorologi Maritim Tanjung Priok

Pada pengolahan model prediksi untuk dataset Stasiun Tanjung Priok, diperlukan pembagian dataset Stasiun Tanjung Priok yang memiliki jumlah data sebanyak 456 data, menjadi 80% untuk data *training* sebesar 364 data dan 20% untuk data *testing* sebesar 92 data. Data *training* ini menggunakan 100 *epochs* untuk permodelan prediksi dan didapatkan akurasi sebesar 87,36 dengan *loss* 32,49. Dari hasil nilai akurasi dan *loss* menunjukkan bahwa model prediksi pada *training* sudah cukup baik dikarenakan nilai akurasi yang tinggi dan mendekati 100.

F. Pengujian Performa Stasiun Meteorologi Kemayoran

Berdasarkan Gambar 3 didapatkan hasil sebagai berikut :

1. Label aman = 151 data terlabel benar (aman) dan 2 data terlabel salah (agak parah).
2. Label agak parah = 8 data terlabel benar (agak parah).
3. Label parah = 1 data terlabel benar (parah), 7 data terlabel salah (2 data terlabel agak parah dan 5 data terlabel sangat parah).

4. Label sangat parah = 20 data terlabel benar (sangat parah).

Sedangkan berdasarkan *classification report*, seperti pada Tabel 4 didapatkan hasil data prediksi probabilitas pada L0 prob didapatkan *precision* = 1,00; *recall* = 0,99; dan *f1-score* = 0,99. Untuk L1 prob didapatkan *precision* = 0,67; *recall* = 1,00; dan *f1-score* = 0,80. Untuk L2 prob didapatkan *precision* = 1,00; *recall* = 0,12; dan *f1-score* = 0,22. Serta L3 prob didapatkan *precision* = 0,80; *recall* = 1,00; dan *f1-score* = 0,89. Perolehan hasil pengujian ini menunjukkan bahwa ketepatan model yang dibuat terhadap prediksi probabilitas yang dihasilkan sudah cukup baik dengan nilai akurasi sebesar 95,24.

G. Pengujian Performa Stasiun Meteorologi Halim Perdana Kusuma

Berdasarkan Gambar 4 didapatkan hasil sebagai berikut : (1)Label aman = 175 data terlabel benar (aman);(2)Label agak parah = 14 data terlabel benar (agak parah), 7 data terlabel salah (3 data terlabel agak parah, 1 data terlabel parah, 3 data terlabel sangat parah);(3)Label parah = 1 data terlabel benar (parah), 16 data terlabel salah (9 data terlabel aman dan 7 data terlabel sangat parah);(4)Label sangat parah = 20 data terlabel benar (sangat parah) dan 3 data terlabel salah (aman).

Sedangkan berdasarkan *classification report*, seperti pada Tabel 5 didapatkan hasil data prediksi probabilitas pada L0 prob didapatkan *precision* = 0,87; *recall* = 1,00; dan *f1-score* = 0,93. Untuk L1 prob didapatkan *precision* = 1,00; *recall* = 0,14; dan *f1-score* = 0,25. Untuk L2 prob didapatkan *precision* = 0,50; *recall* = 0,06; dan *f1-score* = 0,11. Serta L3 prob didapatkan *precision* = 0,67; *recall* = 0,87; dan *f1-score* = 0,75. Perolehan hasil pengujian ini menunjukkan bahwa ketepatan model yang dibuat terhadap prediksi probabilitas yang dihasilkan sudah cukup baik dengan nilai akurasi sebesar 84,32.

H. Pengujian Performa Stasiun Meteorologi Maritim Tanjung Priok

Berdasarkan Gambar 5 didapatkan hasil sebagai berikut :(1)Label aman = 75 data terlabel benar (aman);(2)Label agak parah = 0 data terlabel benar (agak parah) dan 6 data terlabel salah (aman);(3)Label parah = 0 data terlabel benar (parah) dan 5 data terlabel salah (aman);(4)Label sangat parah = 1 data terlabel benar (sangat parah) dan 5 data terlabel salah (aman).

Sedangkan berdasarkan *classification report*, seperti pada Tabel 6 didapatkan hasil data prediksi probabilitas pada L0 prob didapatkan *precision* = 0,82; *recall* = 1,00; dan *f1-score* = 0,90. Untuk L1 prob didapatkan *precision* = 0,00; *recall* = 0,00; dan *f1-score* = 0,00. Untuk L2 prob didapatkan *precision* = 0,00; *recall* = 0,00; dan *f1-score* = 0,00. Serta L3 prob didapatkan *precision* = 1,00; *recall* = 0,17; dan *f1-score* = 0,29. Perolehan hasil pengujian ini menunjukkan bahwa ketepatan model yang dibuat terhadap prediksi probabilitas yang dihasilkan sudah cukup baik dengan nilai akurasi sebesar 82,61. Contoh Data Prediksi Probabilitas dan Pelabelan Stasiun Meteorologi Kemayoran dapat dilihat pada Tabel 7.

V. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:(1)Model prediksi yang dibuat dapat memprediksi tingkat keparahan di

Provinsi DKI Jakarta dengan akurasi pada Stasiun Meteorologi Kemayoran sebesar 93,77; Stasiun Meteorologi Halim Perdana Kusuma sebesar 89,83; Stasiun Maritim Meteorologi Tanjung Priok sebesar 87,36;(2)Didapatkan hasil data prediksi probabilitas Stasiun Meteorologi Kemayoran sebesar 95,24; Stasiun Meteorologi Halim Perdana Kusuma sebesar 84,32; Stasiun Maritim Meteorologi Tanjung Priok sebesar 82,61;(3)Permodelan *Deep Learning* menggunakan *Artificial Neural Network* merupakan metode yang baik untuk memprediksi kasus demam berdarah di Provinsi DKI Jakarta berdasarkan hasil akurasi model *training* dan model prediksi probabilitas;(4)Keunggulan penelitian ini daripada penelitian terdahulu adalah penelitian ini menggunakan metode pembelajaran yang lebih mahir, serta *output* yang dihasilkan berupa tingkat keparahan

penyebaran kasus demam berdarah disetiap kecamatan yang mudah dibaca oleh khalayak umum;(5)Pelabelan kondisi keparahan hasil prediksi probabilitas sangat dibutuhkan untuk mempermudah pembacaan data tabel.

DAFTAR PUSTAKA

- [1] A. Shifa Ichwani and H. A. Wibawa, "Prediksi angka kejadian demam berdarah dengue (dbd) berdasarkan faktor cuaca menggunakan metode extreme learning machine (studi kasus Kecamatan Tembalang)," *Jurnal IPTEK Media Komunikasi Teknologi*, vol. 23, no. 1, pp. 31–38, 2019, doi: 10.31284/j.iptek.2019.v23i1.
- [2] T. Respati, W. Wanti, and R. D. Nindrea, "Dengue Cases Prediction in Kupang," *Global Medical & Health Communication (GMHC)*, vol. 8, no. 3, pp. 219–225, Dec. 2020, doi: 10.29313/gmhc.v8i3.6727.