

## PREDIKSI PENYEBARAN KASUS DEMAM BERDARAH *DENGUE* (DBD) DENGAN METODE *K-NEAREST NEIGHBOR* DI KABUPATEN MALUKU TENGAH

S. Latuconsina<sup>1</sup> dan M. Y. Matdoan<sup>2</sup>

<sup>1,2</sup>Program Studi Statistika FMIPA Universitas Pattimura

Jl. M. J. Putuhena, Kampus Unpatti-Poka, Ambon, 97233, Indonesia

<sup>1</sup>sitha.latuconsina@gmail.com, <sup>2</sup>keepyahya@gmail.com

### ABSTRACT

Dengue Hemorrhagic Fever (DHF) is a disease caused by infection with the dengue virus. Dengue fever is caused by one of four viral serotypes of the flavivirus genus. This disease is a dangerous disease, so predictions need to be made to find out which areas are prone to dengue cases. Making these predictions aims to be able to immediately identify areas with high to low levels of spread to take appropriate action and prevention, thereby reducing the impact of the dengue epidemic. This study obtained the results that the spread of dengue cases in Central Maluku Regency with details of 4 sub-districts that have the potential to spread high dengue fever, namely Banda, Telutih, Salahutu and Masohi Districts. The sub-districts that have the potential to spread moderate dengue fever are Tehoru, Amahai, Saparua, Nusalaut, Haruku Island, Leihitu, West Leihitu and Teon Nila Serua. Furthermore, the sub-districts that have the potential to spread low DBDB are Teluk Elpaputih, North Seram, North West Seram, North East Seram Kobi and North East Seti Seram.

**Keywords** : Dengue Hemorrhagic Fever, K-Nearest Neighbor, Maluku Tengah

### ABSTRAK

Demam Berdarah Dengue (DBD) merupakan suatu penyakit yang disebabkan oleh infeksi virus dengue. DBD disebabkan oleh salah satu dari empat serotipe virus dari genus *flavivirus*. Penyakit ini merupakan salah satu penyakit berbahaya, sehingga perlu dilakukan prediksi untuk mengetahui wilayah mana saja yang rawan terkena kasus DBD. Melakukan prediksi tersebut bertujuan agar dapat dengan segera untuk mengetahui wilayah yang tingkat penyebarannya tinggi hingga rendah untuk melakukan tindakan dan pencegahan yang sesuai, sehingga mengurangi dampak dari penyakit epidemic DBD tersebut. Penelitian ini diperoleh hasil bahwa penyebaran kasus DBD di Kabupaten Maluku Tengah dengan rincian 4 kecamatan yang berpotensi penyebaran DBD tinggi yaitu Kecamatan Banda, Telutih, Salahutu dan Kota Masohi. Kecamatan yang berpotensi penyebaran DBD sedang yaitu Kecamatan Tehoru, Amahai, Saparua, Nusalaut, Pulau Haruku, Leihitu, Leihitu Barat dan Teon Nila Serua. Selanjutnya kecamatan yang berpotensi penyebaran DBDB rendah yaitu Kecamatan Teluk Elpaputih, Seram Utara, Seram Utara Barat, Seram Utara Timur Kobi dan Seram Utara Timur Seti.

**Kata kunci** : Demam Berdarah Dengue, *K-Nearest Neighbor*, Maluku Tengah

## I. PENDAHULUAN

Demam Berdarah Dengue (DBD) merupakan suatu penyakit yang disebabkan oleh infeksi virus dengue. DBD adalah penyakit akut dengan manifestasi klinis pendarahan yang menimbulkan syok yang berujung kematian[1]. DBD disebabkan oleh salah satu dari empat serotipe virus dari genus *flavivirus*, family *flaviviridae*. Hampir setiap tahun terjadi Kejadian Luar Biasa (KLB) di beberapa daerah pada musim penghujan[2]. Penyakit ini masih menjadi masalah kesehatan masyarakat dan endemis di sebagian kabupaten/kota di Indonesia.

Salah satu faktor risiko penularan DBD merupakan pertumbuhan penduduk perkotaan yang cepat, mobilisasi penduduk karena membaiknya sarana dan prasarana transportasi dan terganggu atau melemahnya pengendalian populasi sehingga memungkinkan terjadinya KLB. Faktor risiko lainnya adalah kemiskinan yang mengakibatkan orang tidak mempunyai kemampuan untuk menyediakan rumah yang layak dan sehat, pasokan air minum dan pembuangan sampah yang benar[3]. Tetapi di lain pihak, DBD juga bisa menyerang penduduk yang lebih makmur terutama yang biasa bepergian[4].

Penyakit epidemik seperti DBD merupakan salah satu penyakit berbahaya, sehingga perlu dilakukan prediksi untuk memprediksi wilayah mana saja yang rawan terkena kasus DBD. Melakukan prediksi tersebut bertujuan agar dapat dengan segera untuk mengetahui wilayah yang tingkat penyebarannya tinggi hingga paling rendah untuk melakukan tindakan dan pencegahan yang sesuai, sehingga dampak dari penyakit epidemic DBD tersebut dapat diminimalisir oleh pihak terkait seperti dinas kesehatan Kabupaten Maluku Tengah. Sebelum dinas kesehatan dan pihak terkait melakukan tindakan pencegahan, seharusnya dilakukan pengutamaan wilayah mana saja yang diprediksi berpotensi terkena wabah epidemic tersebut berdasarkan tingkat jumlah kasus DBD[6]. Untuk membantu melakukan prediksi tersebut, dibutuhkan suatu metode yang dapat dengan tepat menentukan tingkat penderita DBD di wilayah Kabupaten Maluku Tengah dengan akurat, salah satunya dengan menggunakan metode *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan salah satu metode untuk mengambil keputusan menggunakan pembelajaran terawasi dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai[7][8].

## II. METODE PENELITIAN

### 2.1. Tipe Penelitian

Penelitian ini menggunakan penelitian kuantitatif yaitu penelitian yang dimulai dari teori, hipotesis, desain penelitian, memilih objek, mengumpulkan data, memproses data, menganalisis data, dan menuliskan kesimpulan.

### 2.2. Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data faktor-faktor yang mempengaruhi penyebaran penyakit Demam Berdarah Dengue (BDB) di Kabupaten Maluku Tengah, adapun

jenis data yang digunakan adalah Data Jumlah Kasus DBD di Kabupaten Maluku Tengah, Data Curah Hujan di Kabupaten Maluku Tengah, Data Kepadatan Penduduk di Kabupaten Maluku Tengah, dan Data Kelembapan Udara di Kabupaten Maluku Tengah. Data-data ini merupakan data sekunder, karena diperoleh dari Dinas Kesehatan Kabupaten Maluku Tengah, dan Badan Pusat Statistik Kabupaten Maluku Tengah.

### 2.3. Analisa Data

Analisa data pada penelitian ini berdasarkan metode *K-Nearest Neighbor* (KNN). metode *K-Nearest Neighbor* (KNN). merupakan sebuah metode *supervised* yang berarti membutuhkan data *training* untuk mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja *K-Nearest Neighbor* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan *k* tetangga (*neighbor*) dalam data pelatihan[9].

Evaluasi dari hasil prediksi penyebaran kasus DBD dilakukan dengan *confusion matrix*. Metode ini merepresentasikan hasil klasifikasi menggunakan matriks yang dapat dilihat pada Tabel 1.

Tabel 1 : Confusion Matrix

Correct Classification	Clasified as		
		+	-
+		True Positive	False Positive
-		False Positive	True Positive

Berdasarkan Tabel 1, dapat dijelaskan bahwa *True Positive* merupakan jumlah *record* positif yang berhasil diklasifikasikan sebagai positif. Sedangkan *false positive* merupakan *record* positif yang salah diklasifikasikan menjadi negatif. Sedangkan *false negative* merupakan *record* negatif yang salah diklasifikasikan sebagai positif, dan untuk *true negative* adalah *record* negatif yang berhasil diklasifikasikan sebagai *record* negatif. Metode pengujian *confusion matrix* dapat menghasilkan perhitungan dengan 4 output, di antaranya yaitu:

$$\begin{aligned}
 \textit{Precision} &= \frac{TP}{TP + FP} \times 100\% \\
 \textit{Recall} &= \frac{TP}{TP + FN} \times 100\% \\
 \textit{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 \textit{Error} &= \frac{FP + FN}{TP + TN + FP + FN} \times 100\%
 \end{aligned}$$

### III. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan dalam penelitian ini berupa hasil nilai akurasi dan prediksi yang dihasilkan dari model untuk mengklasifikasikan data yang digunakan untuk memprediksi daerah dengan kasus demam berdarah di kabupaten Maluku Tengah menggunakan algoritma *K-Nearest Neighbor*.

Setelah pengumpulan data, selanjutnya dilakukan pemahaman terhadap data dengan mendeskripsikan masing-masing data. Deskripsi data ini dilakukan untuk memberi gambaran dari setiap data untuk mengetahui karakteristik data dengan mendeskripsikan tipe dan jenis setiap data yang telah dikumpulkan. Berikut pada Tabel 2 merupakan deskripsi dari masing-masing data.

Tabel 2 : Variabel Penelitian

Data	Tipe Data	Keterangan
Kecamatan	Karakter	Prediktor
Curah Hujan	Numerik	Prediktor
Kepadatan Penduduk	Numerik	Prediktor
Kelembapan	Numerik	Preditor
Demam Berdarah	Kategorikal	Class

Setelah pendeskripsian data, selanjutnya dilakukan pengecekan *Missing Value*. Karena pada data tidak ada data yang hilang, maka dilanjutkan dengan pengecekan outlier dengan mengkonversi nilai data dengan cara standardized (Z-score).

$$Z = \frac{(x - \mu)}{\sigma}$$

Data dikatakan outlier apabila nilai Z yang didapat lebih besar dari 2,5 atau lebih kecil dari -2,5. Dapat dilihat pada Tabel 3.

Tabel 3 : Outlier Data

Kepadatan penduduk	Rata-rata kelembapan	Rata-rata curah hujan
32	84,3	207,325
37	87,25	272,853
37	85,91	195,958
32,2	85,75	222,675
36	91,16	373,133
37	86,58	314,483

Berdasarkan Tabel 3, terlihat bahwa tidak ada data yang outlier. Selanjutnya adalah mentransformasikan data dengan teknik normalisasi data. Normalisasi data dilakukan menggunakan teknik Z-Transformation yaitu sebuah metode matematis yang diperlukan untuk mengurangi

perhitungan dalam sebuah operasi serta mempersempit selisih dari setiap data. Transformasi-Z didefinisikan sebagai :

$$X_{(z)} = \sum_n^{\infty} x(n)z^{-n}$$

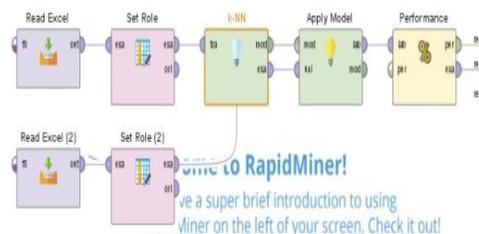
Tabel 4 : Standarisasi Data

No	Kepadatan penduduk	Rata-rata kelembaban	Rata-rata curah hujan	Kasus DBD
1.	-1,315	-1,079	-0,822	Tinggi
2.	0,740	0,182	0,122	Rendah
3.	0,740	-0,391	-0,986	Sedang
4.	-1,233	-0,459	-0,601	Rendah
5.	0,329	1,852	1,567	Rendah
6.	0,740	-0,105	0,722	Tinggi

Setelah melewati proses normalisasi berikut pada Tabel 4 merupakan hasil dari data transformasi yang sudah diproses pada Tabel 1. Selanjutnya adalah tahap modelling, dilakukan pembagian terhadap dataset secara acak menjadi 2 bagian yaitu data training dan data testing dengan rasio perbandingan (80;20). Data training digunakan untuk menghasilkan model prediksi menggunakan algoritma KNN dan data testing digunakan untuk melihat *performance* model prediksi yang dihasilkan. Berikut merupakan hasil pembagian terhadap dataset.

### 1) Pengujian Data dengan Algoritma KNN

Pengolahan data dilakukan dengan menggunakan tools *Rapidminer* yang dimana Algoritma KNN dihitung secara otomatis menggunakan tools tersebut. Pada tahap ini dilakukan 5 kali pengujian dengan nilai K masing-masing K=3, K=5, K=7, K=9 dan K=11. Proses K-Fold Cross Validation menggunakan skema 10-Fold Cross Validation, dimana dataset akan dibagi menjadi N bagian secara acak. Fold ke-1 adalah Ketika bagian ke-1 menjadi data testing dan sisanya menjadi data training, demikian seterusnya hingga sampai fold ke-10 bagian ke-10. Proses K-Fold Cross Validation ditambahkan pada pengujian dengan tujuan untuk meningkatkan performance dari Algoritma KNN. Berikut pada gambar 4 merupakan rangkaian proses pada tools Rapidminer.



Gambar 1 : Olahan RapidMiner

## 2) Evaluasi

Tahap evaluasi dilakukan dengan melihat *performance* kinerja klasifikasi yang telah dilakukan pada pemodelan sebelumnya. menggunakan *confusion matrix* untuk melihat nilai *accuracy*, *recall* dan *precision*. Berikut pada Tabel 5 – 7 merupakan hasil *confusion matrix* dari Algoritma KNN dengan nilai K=3, K=4 dan K=5.

Tabel 5 : Hasil *Confusion Matrix* K=3

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	12	3	2	80,00%
Pred Sedang	1	4	1	50,00%
Pred Rendah	1	4	4	65,00%
Class Recall	72%	60,14%	55%	

Tabel 6 : Hasil *Confusion Matrix* K=4

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	12	3	1	75,00%
Pred Sedang	0	4	0	40,00%
Pred Rendah	2	4	5	70,00%
Class Recall	72%	60,14%	50%	

Tabel 7 : Hasil *Confusion Matrix* K=5

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	14	3	0	80,00%
Pred Sedang	0	3	0	36,00%
Pred Rendah	0	3	4	55,00%
Class Recall	72%	57%	40%	

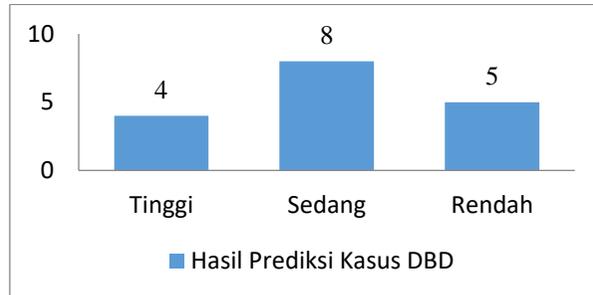
Berdasarkan Tabel 7 confusion matrix diperoleh nilai *class recall* dan *class precision*. Berikut pada Tabel 8 merupakan nilai *accuracy* dari pengujian data.

Tabel 8 : Hasil Accuracy

KNN	Nilai K		
	K=3	K=4	K=5
Accuracy	75,12%	72%	68,21%

Berdasarkan hasil accuracy diatas dapat dilihat nilai *accuracy* K=3 merupakan *accuracy* tertinggi dengan menggunakan Algoritma KNN dan 10-Fold *Cross Validation* mendapatkan nilai *accuracy* sebesar 75.12%.

Dalam proses data mining untuk klasifikasi kasus demam berdarah di Kabupaten Maluku Tengah. Algoritma *K-Nearest Neighbor* digunakan untuk mencari nilai akurasi yang menghasilkan prediksi kasus demam berdarah pada setiap kecamatan di kabupaten Maluku tengah. Berikut pada Gambar 2 merupakan diagram prediksi yang telah dihasilkan.



Gambar 2 : Hasil Prediksi

Berdasarkan hasil prediksi tersebut, terdapat rincian 4 kecamatan yang berpotensi penyebaran DBD tinggi yaitu Kecamatan Banda, Telutih, Salahutu dan Kota Masohi. Kecamatan yang berpotensi penyebaran DBD sedang yaitu Kecamatan Tehoru, Amahai, Saparua, Nusalaut, Pulau Haruku, Leihitu, Leihitu Barat dan Teon Nila Serua. Selanjutnya kecamatan yang berpotensi penyebaran DBD rendah yaitu Kecamatan Teluk Elpapatih, Seram Utara, Seram Utara Barat, Seram Utara Timur Kobi dan Seram Utara Timur Seti.

#### IV. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan yaitu mengenai jumlah prediksi kasus DBD di Kabupaten Maluku Tengah berdasarkan data pada tahun 2016-2021 dengan menggunakan algoritma *K-Nearest Neighbor* dan metode *10-fold cross validation* menghasilkan *performance* terbaik menggunakan nilai  $K = 3$  dan nilai *accuracy* sebesar 75,27%. Secara keseluruhan, penerapan Algoritma *K-Nearest Neighbor* dapat memberikan *accuracy* dan prediksi yang cukup baik. Adapun kecamatan yang diprediksi terjadi penyebaran kasus demam berdarah dengan rincian 4 kecamatan yang berpotensi penyebaran DBD tinggi yaitu Kecamatan Banda, Telutih, Salahutu dan Kota Masohi. Kecamatan yang berpotensi penyebaran DBD sedang yaitu Kecamatan Tehoru, Amahai, Saparua, Nusalaut, Pulau Haruku, Leihitu, Leihitu Barat dan Teon Nila Serua. Selanjutnya kecamatan yang berpotensi penyebaran DBD rendah yaitu Kecamatan Teluk Elpapatih, Seram Utara, Seram Utara Barat, Seram Utara Timur Kobi dan Seram Utara Timur Seti.

## DAFTAR PUSTAKA

- [1] Ariati Jusniar, A. A., 2014. Model Prediksi Demam Berdarah Dengue (DBD) Berdasarkan Faktor Iklim di Kota Bogor, Jawa Tengah. *Buletin Penelitian Kesehatan*, 42(4), pp. 249-256.
- [2] Aryu, C., 2010. Demam Berdarah Dengue : Epidemiologi, Patogenesis, dan Faktor Risiko Penularan. *Aspirator*, 2(110), pp. 110-119.
- [3] Didik, P., 2016. *Pengaruh Pendidikan Kesehatan Terhadap Pengetahuan dan Sikap Tentang Pencegahan DBD di Wilayah Puskesmas Purbalingga Kabupaten Purbalingga*, Purwokerto: Fakultas Ilmu Kesehatan UMP.
- [4] Dinkes, 2021. *Data Demam Berdarah Dengue*, Masohi: Dinas Kesehatan Maluku Tengah.
- [5] Indonesia, K. K. R., 2021. *Data DBD Indonesia*, jakarta: GERMAS.
- [6] Johar.T.Ashari, Y. A. K., 2016. Implementasi Metode K-Nearest Neighbor (KNN) Dan Sample Additive Weighting (SAW) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka. *Pseudocode*, III(101), pp. 98-112.
- [7] Sukohar, A., 2014. Demam Berdarah Dengue (DBD). *Fakultas Kedokteran Universitas Lampung*, 2(2), pp. 1-15.
- [8] WHO, 1999. *Panduan Lengkap Pencegahan & Pengendalian Dengue & DBD*. - ed. Jakarta: EGC.
- [9] Widoyono, 2008. *Penyakit Tropis Epidemiologi, Penularan, Pencegahan, dan Pemberantasannya*. - ed. Jakarta: Erlangga.