



Penerapan Metode *K-Nearest Neighbor* untuk Mengklasifikasi Penyebaran Kasus Demam Berdarah *Dengue* (DBD) di Kabupaten Maluku Tenggara

Muhammad Yahya Matdoan
Program Studi Statistika FMIPA Universitas Pattimura
Email: keepyahya@gmail.com

ABSTRAK

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan masuk ke peredaran darah manusia. Penyakit ini merupakan penyakit berbahaya yang sering menimbulkan kekhawatiran masyarakat karena perjalanan penyakitnya cepat dan dapat menyebabkan kematian dalam waktu singkat. Oleh karena itu, perlu dilakukan kajian tentang penyebarannya sehingga dapat diambil tindakan cepat dalam mencegah kasus tersebut, Salah satu metode yang dapat digunakan yaitu metode *K-Nearest Neighbor* (KNN). K-NN merupakan suatu bentuk model pendukung keputusan yang dapat mengklasifikasikan data berdasarkan jarak terdekat. Data yang digunakan dalam penelitian ini bersumber dari BPS Kabupaten Maluku Tenggara tahun 2021. Penelitian ini diperoleh hasil bahwa terdapat 2 kelompok penyebaran DBD di Kabupaten Maluku Tenggara yaitu Kecamatan yang berpotensi DBD tinggi yaitu terdiri dari Kecamatan Kei Besar, Kecamatan Kei Kecil dan Kecamatan Kei Besar Selatan Barat. Selanjutnya kecamatan yang berpotensi penyebaran DBD rendah yaitu Kecamatan Kei Besar Utara Barat, Kecamatan Kei Besar Selatan, Kecamatan Kei Besar Timur Selatan, Kecamatan Kei Besar Utara Timur, Kecamatan Hoat Sorbay, Kecamatan Kei Kecil Barat, Kecamatan Kei Kecil Timur dan Kecamatan Manyeuw.

Kata Kunci: Demam Berdarah Dengue, *K-Nearest Neighbor*, Maluku Tenggara.

ABSTRACT

Dengue Hemorrhagic Fever (DHF) is a disease caused by the dengue virus which is transmitted through the bite of the Aedes Aegypti mosquito and enters human blood circulation. This disease is a dangerous disease that often causes public concern because the course of the disease is fast and can cause death in a short time. Therefore, it is necessary to study its spread so that prompt action can be taken to prevent this case. One method that can be used is the K-Nearest Neighbor (KNN) method. K-NN is a form of decision support model that can classify data based on the shortest distance. The data used in this study came from the BPS of Southeast Maluku Regency in 2021. This research showed that there are 2 groups of DHF spread in Southeast Maluku Regency, namely Districts with high potential for DHF, which consist of Kei Besar District, Kei Kecil District and Kei Besar Selatan District. West. Furthermore, the sub-districts that have the potential to spread low DHF are the North West Kei Besar District, South Kei Besar District, South East Kei Besar District, East North Kei Besar District, Hoat Sorbay District, West Kei Kecil District, East Kei Kecil District and Manyeuw District.

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

1. PENDAHULUAN

Penyakit menular merupakan masalah kesehatan yang paling serius di Indonesia bahkan di dunia. Penyakit menular adalah suatu penyakit yang disebabkan oleh mikroorganisme baik bakteri, virus, maupun jamur, yang bisa ditularkan dari satu orang penderita kepada orang sehat sehingga menyebabkan sakit seperti sumber penularan, salah satu penyakitnya yaitu Demam Berdarah Dengue (DBD). DBD merupakan penyakit yang disebabkan oleh virus *Dengue* yang ditularkan melalui gigitan nyamuk *Aedes Aegypti* dan masuk ke peredaran darah manusia. Penyakit ini merupakan penyakit berbahaya yang sering menimbulkan kekhawatiran masyarakat karena perjalanan penyakitnya cepat dan dapat menyebabkan kematian dalam waktu singkat (Depkes RI, 2005).

Menurut Supari (2019) penyebab meningkatnya jumlah kasus dan semakin bertambahnya wilayah yang terjangkit, antara lain karena semakin padatnya penduduk dan mobilitas penduduk. Selain itu semakin baiknya transportasi dari suatu daerah ke daerah lainnya serta adanya pemukiman-pemukiman baru juga menjadi penyebab meningkatnya kasus DBD. Faktor lain yang mempengaruhi terjadinya DBD adalah penyimpangan pola hujan, faktor musim, perilaku masyarakat menyimpan air secara tradisional, kurang partisipasi masyarakat dalam Pemberantasan Sarang Nyamuk (PSN), kurang pengetahuan masyarakat tentang gejala DBD dan keterlambatan membawa penderita ke tempat pelayanan kesehatan.

Penyakit epidemik seperti DBD merupakan salah satu penyakit berbahaya, sehingga perlu dilakukan prediksi untuk memprediksi wilayah mana saja yang rawan terkena kasus DBD. Melakukan klasifikasi 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 Tenggara. Oleh karena itu, dibutuhkan suatu metode yang dapat menentukan klasifikasi penderita DBD di wilayah Kabupaten Maluku Tenggara dengan akurat, salah satunya dengan menggunakan metode *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut. Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap vektor data pembelajaran dihitung, dan diambil sejumlah K yang paling mendekati. Titik yang baru klasifikasinya diprediksi termasuk pada klasifikasi terbanyak dari titik-titik tersebut. Feldman, R. d. (2007).

1.1. Data Mining

Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat (Sikumbang, E. D, 2018). Data mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi (Septiani, W. D, 2017). Proses dalam tahap data mining terdiri dari tiga langkah utama, yaitu data preparation pada langkah ini, data dipilih, dibersihkan, dan dilakukan preprocessed mengikuti pedoman dan *knowledge* dari ahli domain

yang menangkap dan mengintegrasikan data internal dan eksternal kedalam tinjauan organisasi secara menyeluruh (Widaningsih, S. (2019). Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai. Namun semakin besar data yang diolah maka semakin besar pula waktu dan prosesnya (Mujiasih, S. (2011).

1.2. *K-Nearest Neighbor*

K-Nearest Neighbor (K-NN) adalah salah satu metode dimana metode ini melakukan klasifikasi berdasarkan data training atau data pembelajaran dilihat dari jarak yang paling dekat dengan objek berdasarkan nilai k. Tujuan dari metode ini adalah mencari hasil klasifikasi berdasarkan jarak terdekat dari setiap objek. Untuk menentukan jarak terdekat tersebut terlebih dahulu data dibagi menjadi data training dan data testing, setelah diperoleh data training dan data testing kemudian di hitung jarak masing-masing data testing (*euclidean distance*) terhadap data training. Rumus yang digunakan untuk menghitung jarak bisa dilihat pada Persamaan 1.

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Keterangan:

$d(x_i, x_j)$: Euclidean Distance (Jarak Euclidean)

x_i, x_j : record ke-m, record ke-n

(a_r) : data ke-r

n : dimensi objek

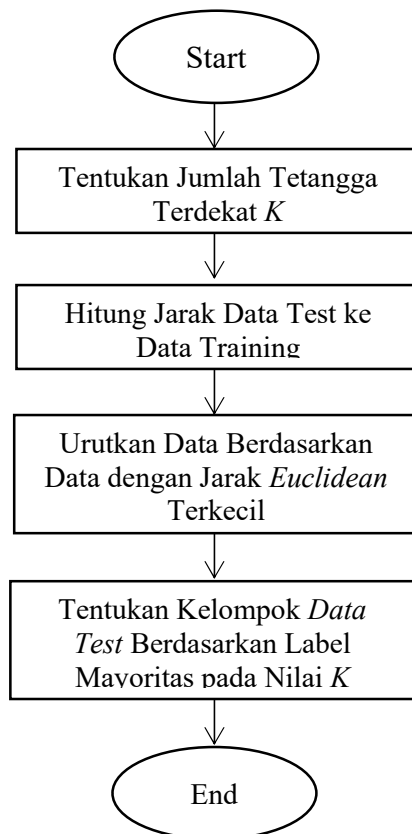
2. METODE

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Maluku Tenggara tahun 2021. Data yang digunakan dalam penelitian ini berupa data Data Jumlah Kasus DBD di Kabupaten Maluku Tenggara, Curah Hujan, Data Kepadatan Penduduk dan Data Kelembapan Udara di Kabupaten Maluku Tenggara. Penelitian ini menggunakan metode KNN dengan tahapan pada Gambar 1.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

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 Tenggara menggunakan algoritma *K-Nearest Neighbor*.



Gambar 1. Prosedur Algoritma KNN

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. Tabel 1 merupakan deskripsi dari masing-masing data.

Tabel 1. 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 *standardzed* (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. Outlier Data dapat dilihat pada Tabel 2.

Tabel 2. *Outlier Data*

Kepadatan Penduduk	Rata-Rata Kelembapan	Rata-Rata Curah Hujan
22	74,12	307,314
27	77,25	272,452
26	75,91	295,558
28	75,75	321,275
26	81,26	323,233
25	76,38	115,383

Berdasarkan Tabel 2, 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 3. Standarisasi Data

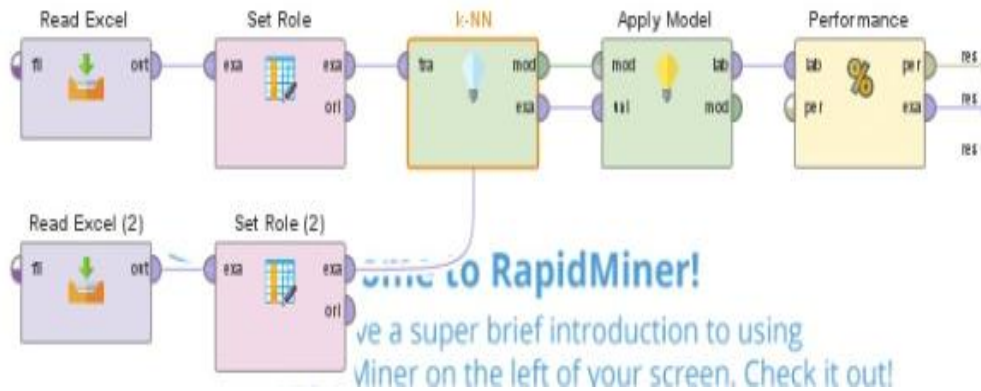
No	Kepadatan Penduduk	Rata-Rata Kelembapan	Rata-Rata Curah Hujan	Kasus DBD
1.	1,425	-1,129	1,122	Tinggi
2.	1,120	0,108	0,132	Rendah
3.	1,350	0,121	0,126	Sedang
4.	-1,533	-0,559	0,501	Rendah
5.	0,621	1,152	1,527	Rendah
6.	0,140	0,205	1,212	Tinggi

Setelah melewati proses normalisasi, Tabel 3 merupakan hasil dari data transformasi yang sudah diproses. 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.

3.2. 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=2, K=4, K=6, K=8 dan K=10. 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 2 merupakan rangkaian proses pada tools Rapidminer.



Gambar 2. Olahan RapidMiner

3.3. 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 4 s.d. Tabel 6 merupakan hasil *confusion matrix* dari Algoritma KNN dengan nilai K=2, K=3, dan K=4.

Tabel 4. Hasil *Confusion Matrix* K=2

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	12	3	2	75,00%
Pred Sedang	1	4	1	50,00%
Pred Rendah	1	4	4	58,00%
Class Recall	81%	70,32%	65%	

Tabel 5. Hasil *Confusion Matrix* K=3

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	12	3	1	72,00%
Pred Sedang	0	4	0	40,00%
Pred Rendah	2	4	5	70,00%
Class Recall	81%	70,32%	60%	

Tabel 6. Hasil *Confusion Matrix* K=4

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred Tinggi	14	3	0	75,00%
Pred Sedang	0	3	0	36,00%
Pred Rendah	0	3	4	55,00%
Class Recall	81%	67%	55%	

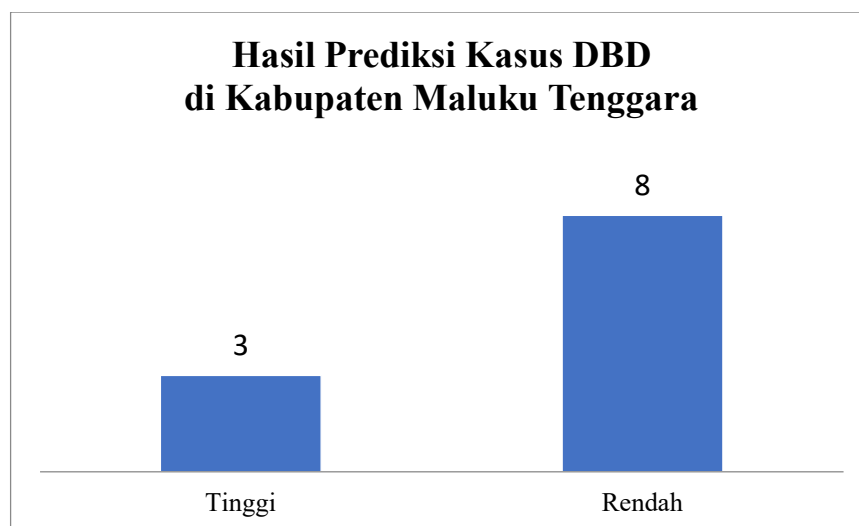
Berdasarkan Tabel confusion matrix diperoleh nilai *class recall* dan *class precision*. Berikut pada Tabel 7 merupakan nilai *accuracy* dari pengujian data.

Tabel 7. Hasil *Accuracy*

KNN	Nilai K		
	K=2	K=3	K=4
Accuracy	72,31%	70,17%	70,05%

Berdasarkan hasil *accuracy* diatas dapat dilihat nilai *accuracy* $K=2$ merupakan *accuracy* tertinggi dengan menggunakan Algoritma KNN dan 10-Fold *Cross Validation* mendapatkan nilai *accuracy* sebesar 72,31%.

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



Gambar 2. Hasil Prediksi

Berdasarkan hasil prediksi tersebut, terdapat rincian 3 kecamatan yang berpotensi penyebaran DBD tinggi yaitu Kecamatan Kei Besar, Kecamatan Kei Kecil dan Kecamatan Kei Besar Selatan Barat. Kecamatan yang berpotensi penyebaran DBD rendah yaitu Kecamatan Kei Besar Utara Barat, Kecamatan Kei Besar Selatan, Kecamatan Kei Besar Timur Selatan, Kecamatan Kei Besar Utara Timur, Kecamatan Hoat Sorbay, Kecamatan Kei Kecil Barat, Kecamatan Kei Kecil Timur dan Kecamatan Manyeuw.

4. SIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan yaitu mengenai jumlah prediksi kasus DBD di Kabupaten Maluku Tenggara berdasarkan data pada tahun 2021 dengan menggunakan algoritma *K-Nearest Neighbor* dan metode *10-fold cross validation* menghasilkan *performance* terbaik menggunakan nilai $K = 2$ dan nilai *accuracy* sebesar 72,31%. 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 3 kecamatan yang berpotensi penyebaran

DBD tinggi yaitu Kecamatan Kei Besar, Kecamatan Kei Kecil dan Kecamatan Kei Besar Selatan Barat. Kecamatan yang berpotensi penyebaran DBD rendah yaitu Kecamatan Kei Besar Utara Barat, Kecamatan Kei Besar Selatan, Kecamatan Kei Besar Timur Selatan, Kecamatan Kei Besar Utara Timur, Kecamatan Hoat Sorbay, Kecamatan Kei Kecil Barat, Kecamatan Kei Kecil Timur dan Kecamatan Manyeuw.

5. DAFTAR PUSTAKA

- Departemen Kesehatan RI. (2005). Pencegahan dan Pemberantasan Demam Berdarah Dengue di Indonesia. Official Website Departemen Kesehatan Republik Indonesia. <http://www.depkes.go.id> (12 Agustus 2015). Departemen Kesehatan RI.(2007). Penyelidikan dan Penanggulangan KLB. *Jurnal Kesehatan Lingkungan Universitas Airlangga Vol, 8(2)*, 172-183.
- Ginanjari, G. (2008). *Demam berdarah*. PT Mizan Publika.
- Kang, S. (2021). K-nearest neighbor learning with graph neural networks. *Mathematics*, 9(8), 830.
- Liantoni, F. (2015). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *Ultimatics: Jurnal Teknik Informatika*, 7(2), 98-104.
- Lidya, S. K. (2014). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) Dan K-Nearest Neighbor (K-NN).
- Mujiasih, S. (2011). Pemanfaatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi dan Geofisika*, 12(2).
- Rinawan, F. (2015). Dampak perubahan iklim terhadap kejadian demam berdarah di Jawa-Barat. *Jurnal Sistem Kesehatan*, 1(1).
- Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1).
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76-84.
- Shaul, H., Feldman, D., & Rus, D. (2018). Secure \$ k \$-ish Nearest Neighbors Classifier. *arXiv preprint arXiv:1801.07301*.
- Sikumbang, E. D. (2018). Penerapan data mining penjualan sepatu menggunakan metode algoritma apriori. *Jurnal Teknik Komputer AMIK BSI*, 4(1), 156-161.
- Sintorini, M. M. (2007). Pengaruh iklim terhadap kasus demam berdarah dengue. *Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal)*, 2(1), 11-18.
- Supari, S. (2019). Desain Generator Ozone dengan Teknologi Plasma DBD (Die-lectric Barrier Discharge). *Elektrika*, 11(1), 1-8.
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naive Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16-25.
- Wirayoga, M. A. (2013). Hubungan kejadian demam berdarah dengue dengan Iklim di Kota Semarang tahun 2006-2011. *Unnes Journal of Public Health*, 2(4).