

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

***PREDICTION OF THE SPREAD OF DENGUE HEMORRHAGIC FEVER (DHF)
USING K-NEAREST NEIGHBOR METHOD IN KARAWANG DISTRICT***

Bimo Aditya Wahyudi¹, Hannie², Aries Soeharso³

¹²³Universitas Singaperbangsa Karawang
bimo.aditya17024@student.unsika.ac.id

ABSTRACT

*Dengue Hemorrhagic Fever (DHF) is an epidemic disease. Karawang Regency is one of the endemic areas for Dengue Hemorrhagic Fever (DHF) in West Java, Indonesia. The emergence of this disease is related to regional conditions such as the environment and weather as well as community behavior which is also related to the spread of Dengue cases. Epidemic diseases such as Dengue is one of the dangerous diseases, so it is necessary to make predictions to predict which areas are prone to Dengue cases. Making these predictions aims to be able to immediately find out which areas have the highest to lowest levels of distribution to take appropriate action and prevention. To help make these predictions, the methods are used *K-Nearest Neighbor (K-NN)* and *K-Fold Cross Validation*. By using data in 2016-2020 using 5 parameter data as input with a total of 150 records. The results of the experiment resulted in a performance with the value accuracy best of 81.27% using the value of $K=5$ and *10-Fold Cross Validation*. Prediction results are shown by mapping the Karawang Regency area by District. A total of 18 sub-districts with a high level of patience, 7 sub-districts with a medium distribution level, and 5 sub-districts with a low distribution level.*

Keywords: *Dengue Hemorrhagic Fever (DHF), K-Nearest Neighbor (K-NN), Classification, Data Mining.*

ABSTRAK

Demam Berdarah *Dengue* (DBD) merupakan salah satu penyakit epidemik. Kabupaten Karawang merupakan salah satu daerah endemis Demam Berdarah *Dengue* (DBD) di Jawa Barat, Indonesia. Munculnya penyakit ini berkaitan dengan kondisi wilayah seperti lingkungan dan juga cuaca serta perilaku masyarakat juga berkaitan terhadap penyebaran kasus DBD. 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 mengetahui wilayah mana yang tingkat persebaran paling tinggi hingga paling rendah untuk melakukan tindakan dan pencegahan yang sesuai. Untuk membantu melakukan prediksi tersebut, digunakan metode *K-Nearest Neighbor (K-NN)* dan *K-Fold Cross Validation*. Dengan menggunakan data pada tahun 2016-2020 menggunakan 5 data parameter sebagai *input* dengan total 150 *records*. Hasil percobaan menghasilkan *performance* dengan nilai *accuracy* terbaik sebesar 81.27% menggunakan nilai $K=5$ dan *10-Fold Cross Validation*. Hasil prediksi ditunjukkan dengan pemetaan wilayah Kabupaten Karawang berdasarkan Kecamatan. Sebanyak 18 Kecamatan dengan tingkat persebaran tinggi, 7 Kecamatan dengan tingkat persebaran sedang, dan 5 Kecamatan dengan tingkat persebaran rendah.

Kata Kunci: Demam Berdarah *Dengue* (DBD), *K-Nearest Neighbor (K-NN)*, Klasifikasi, *Data Mining*.

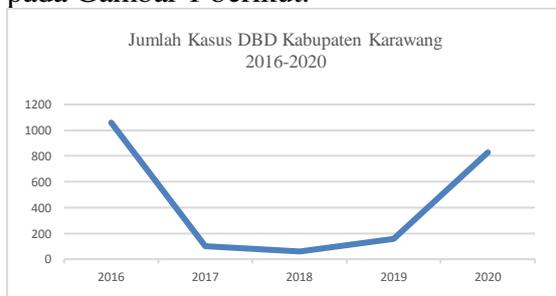
PENDAHULUAN

Luas wilayah Daerah Kabupaten Karawang seluas 1.753,27 Km² atau 3,73% dari luas Provinsi Jawa Barat. Secara geografis wilayah Kabupaten

Karawang terletak antara 070-02-1070-40 B dan 50-56-60-34 LS, termasuk daerah dataran yang relatif rendah, dan mempunyai variasi ketinggian wilayah antara 0 - 1.279 meter di atas permukaan

laut. Kabupaten Karawang merupakan salah satu daerah endemis Demam Berdarah *Dengue* (DBD) di Jawa Barat.

Demam Berdarah *Dengue* (DBD) merupakan salah satu penyakit epidemik. Suatu penyakit bisa dikatakan sebagai penyakit epidemik apabila sudah mewabah di suatu wilayah yang terbilang cukup luas, dengan penyebaran yang cukup cepat dan sulit untuk diprediksi. Penyakit DBD merupakan penyakit yang diakibatkan dari gigitan nyamuk *Aedes Aegypti* yang terjangkit dan menularkan virus *Dengue* ke manusia. Munculnya penyakit ini berkaitan dengan kondisi wilayah seperti lingkungan dan juga cuaca serta perilaku masyarakat juga berkaitan terhadap penyebaran kasus DBD. Kurangnya pengetahuan warga mengenai perilaku hidup bersih dan sehat serta kurangnya pengetahuan warga tentang pencegahan penyakit DBD juga menjadi salah satu faktor penyebab terjadinya penyakit DBD. Banyaknya kasus yang terjadi selama 5 tahun kebelakang dapat dilihat pada Gambar 1 berikut.



Gambar 1. Grafik Kasus DBD 2016-2020
Sumber: Dinas Kesehatan Kabupaten Karawang

Berdasarkan catatan Dinas Kesehatan Kabupaten Karawang pada tahun 2016 terdapat 1059 kasus, tahun 2017 terdapat 102 kasus, 2018 sebanyak 60 kasus, dan 2019 ada 159 kasus. Demam Berdarah *Dengue* (DBD) bahkan tercatat pada tahun 2020 sampai bulan Juni sudah mencapai 827 kasus di wilayah Kabupaten Karawang. Diantara banyaknya kasus DBD tersebut tentu ada beberapa yang berujung dengan

kematian. Penyebab kematian tersebut dipicu oleh beberapa faktor, antara lain keterlambatan penanganan medis dan rendahnya kesadaran masyarakat akan pencegahannya. Hal ini patut menjadi isu yang wajib diwaspadai oleh dinas kesehatan, khususnya dinas kesehatan Karawang.

Penelitian sebelumnya menyebutkan maraknya kasus DBD saat pergantian musim hujan ke musim kemarau, hal ini menandakan iklim di Kabupaten Karawang merupakan salah satu faktor penyebab terjadinya penyebaran kasus DBD (Indriani et al., 2013). Penelitian lainnya juga memaparkan Kepadatan Penduduk, laju pertumbuhan mobilitas penduduk, dan curah hujan, merupakan faktor yang paling berpengaruh dan prediksi dengan menggunakan Algoritma K-NN memberikan hasil yang cukup baik (Salam, Prasetyowati, & Sibaroni, 2021).

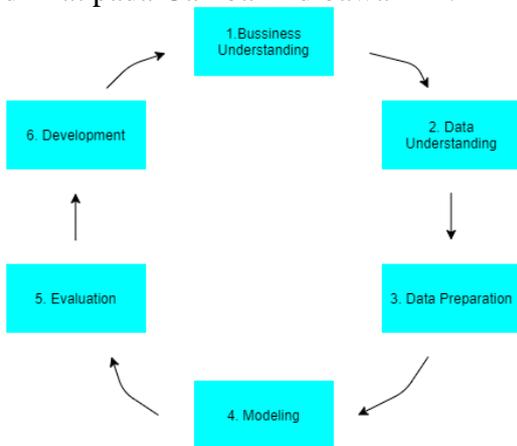
Penyakit epidemik seperti demam berdarah *Dengue* merupakan salah satu penyakit berbahaya, sehingga perlu dilakukan prediksi untuk memprediksi wilayah mana saja yang rawan terkena kasus DBD. Melakukan prediksi tersebut bertujuan agar dapat dengan segera untuk mengetahui wilayah yang tingkat persebaran paling tinggi hingga paling rendah untuk melakukan tindakan dan pencegahan yang sesuai, sehingga dampak dari penyakit epidemik DBD tersebut dapat diminimalisir oleh pihak terkait seperti Dinas Kesehatan Kabupaten Karawang. Sebelum dinas kesehatan dan pihak terkait melakukan tindakan pencegahan, seharusnya dilakukan pengutamaan wilayah mana saja yang diprediksi berpotensi terkena wabah epidemik tersebut berdasarkan tingkat jumlah kasus DBD.

Untuk membantu melakukan prediksi tersebut, dibutuhkan suatu metode yang dapat dengan tepat menentukan tingkat penderita DBD di wilayah Kabupaten

Karawang dengan akurat, salah satunya dengan menggunakan metode *K-Nearest Neighbor* (K-NN)(Shokrzade, dkk., 2021; Shamrat, dkk., 2021; Hertina, dkk., 2021).

METODE

Metodologi yang digunakan pada penelitian ini menggunakan proses metode Cross-Standard Industry-*Data Mining* (CRISP-DM). Tahapan penelitian ini meliputi pemahaman bisnis, pemahaman data, pengolahan data, pemodelan dan evaluasi. Tahapan-tahapan metode CRISP-DM dapat dilihat pada Gambar 2 dibawah ini.



Gambar 2. Tahapan CRISP-DM

1. Fase Pemahaman Tujuan (*Bussiness Understanding*)

Pada fase *Bussiness Understanding*, tahapan pertama yang dibuat adalah mengelompokkan kondisi kasus penyebaran DBD atau tidak berdasarkan atribut-atribut yang berpengaruh. Lalu mengidentifikasi pola terkait dataset mengenai data kasus DBD, data curah hujan, data kepadatan penduduk, data kelembapan dan data mobilitas penduduk untuk memprediksi daerah penyebaran DBD di Kabupaten Karawang.

2. Fase Pemahaman Data (*Data Understanding*)

Pada fase kedua, bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami aset data yang dimiliki.

Data yang digunakan pada tahun 2016 – 2020. Sumber data yang digunakan pada penelitian ini melalui tahap observasi dengan melakukan pengamatan pada Dinas Kesehatan Kabupaten Karawang dan Badan Pusat Statistik Karawang. Terdapat 5 parameter yang digunakan diantaranya data kasus DBD, data curah hujan, data kepadatan penduduk, data kelembapan dan data mobilitas penduduk.

3. Fase Pengolahan Data (*Data Preparation*)

Pada fase ini, yakni menyiapkan dataset yang akan diolah menggunakan algoritma K-NN. Penyiapan data ini penting terutama untuk menyusun data sesuai format yang sesuai dengan algoritma K-NN. Pada fase ini terdapat beberapa proses, yaitu pemilihan data, pembersihan data dan transformasi data.

4. Fase Pemodelan (*Modelling*)

Pada fase permodelan ini akan digunakan teknik *Data Mining* dengan metode klasifikasi menggunakan algoritma K-NN dengan tools *Rapidminer*, yang nantinya akan menghasilkan aturan klasifikasi dan variabel paling berpengaruh dalam memprediksi daerah penyebaran DBD.

5. Fase Evaluasi (*Evaluation*)

Pada fase ini dilakukan evaluasi terhadap keefektifan dan kualitas model yang digunakan, selanjutnya didapatkan hasil prediksi untuk algoritma K-NN. Selanjutnya divalidasi menggunakan metode *K-Fold Cross Validation* untuk memvalidasi hasil tingkat akurasi.

6. Fase Penyebaran (*Deployment*)

Pada fase ini pola informasi yang dihasilkan dari proses *Data Mining* yang telah dibuat akan dipresentasikan dalam bentuk yang dapat dipahami oleh pihak yang berkaitan atau berhubungan serta

dalam bentuk laporan sebagai acuan dalam memprediksi daerah penyebaran penyakit DBD.

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 Karawang menggunakan algoritma *K-Nearest Neighbor*.

1. Business Understanding

Tahapan Business Understanding digunakan untuk memahami tujuan yang ingin dicapai dilihat dari perspektif bisnis. beberapa tahapan untuk pemahaman latar belakang dan tujuan diantaranya menentukan tujuan bisnis, menilai situasi dan menentukan tujuan *Data Mining*.

a. Menentukan Tujuan Bisnis

Berdasarkan permasalahan yaitu dengan tingginya kasus demam berdarah yang terjadi di beberapa wilayah pada Kabupaten Karawang. Maka dilakukan tujuan untuk memprediksi daerah yang rawan terjangkit demam berdarah di Kabupaten Karawang dengan mengklasifikasikan data menggunakan teknik pemodelan. Sehingga nantinya dapat membantu instansi terkait sebagai referensi dalam mengantisipasi daerah yang rawan demam berdarah di Kabupaten Karawang dengan meninjau hasil dari data yang telah diolah sebelumnya.

b. Menilai Situasi

Beberapa faktor penyebab demam berdarah yang menjadi informasi dalam penelitian ini. Faktor penyebab demam berdarah yang digunakan pada penelitian ini diantaranya rata-rata curah hujan, kelembapan, kepadatan penduduk, mobilitas penduduk dan kasus demam berdarah yang terjadi di Kabupaten Karawang.

c. Menentukan Tujuan *Data Mining*

Tujuan *Data Mining* pada penelitian ini adalah menggali pengetahuan baru (discovering knowledge) tentang pola (pattern) mengenai daerah penyebaran demam berdarah di Kabupaten Karawang menggunakan algoritma *K-Nearest Neighbor*. Untuk mendapatkan hasil model terbaik dalam memprediksi, maka menggunakan metode evaluasi dan validasi diantaranya *K-Fold Cross Validation* untuk mengevaluasi kinerja dari algoritma KNN dan confusion matrix untuk mengetahui hasil akurasi.

2. Data Understanding

Tahapan *Data Understanding* digunakan untuk mengenal dan memahami data yang akan diteliti. Proses data understanding ini dilakukan dengan beberapa tahapan untuk pengenalan terhadap data diantaranya proses pengumpulan data dan deskripsi data.

a. Pengumpulan Data

Pada tahap pengumpulan data, data yang digunakan pada penelitian ini bersifat sekunder, dimana data diperoleh melalui situs web Badan Pusat Statistik (BPS) Kabupaten Karawang (<https://karawangkab.bps.go.id/>) dan Dinas Kesehatan Kabupaten Karawang (<https://dinkes.karawangkab.go.id/>).

Terdapat beberapa data yang telah dikumpulkan pada penelitian ini yaitu kecamatan, rata-rata curah hujan, kepadatan penduduk, kelembapan, mobilitas penduduk dan data kasus demam berdarah. Data yang dikumpulkan pada tahun 2016-2020, dimana terdiri dari 30 kecamatan di Kabupaten Karawang.

b. Representasi Pengetahuan

Setelah tahap pengumpulan data, selanjutnya dilakukan pemahaman

terhadap data dengan mendeskripsikan masing-masing data. Deskripsi data ini dilakukan untuk memberikan gambaran dari setiap data untuk mengetahui karakteristik data dengan mendiskripsikan tipe dan jenis pada setiap data yang telah dikumpulkan. Berikut pada Tabel 1 merupakan deskripsi dari masing-masing data.

Tabel 1. Deskripsi Data

Data	Tipe Data	Keterangan
Kecamatan	Karakter	<i>Prediktor</i>
Curah Hujan	Numerik	<i>Prediktor</i>
Kepadatan Penduduk	Numerik	<i>Prediktor</i>
Kelembapan	Numerik	<i>Prediktor</i>
Mobilitas Penduduk	Numerik	<i>Prediktor</i>
Demam Berdarah	Kategorikal	<i>Class</i>

3. Data Preparation

Pada tahap data preparation ini mencakup semua kegiatan dalam menyiapkan dan membangun dataset yang akan diterapkan ke dalam tools pemodelan untuk selanjutnya dilakukan proses *Data Mining*. Dataset yang digunakan pada penelitian ini sebanyak 150 *record* data dengan menggunakan 5 parameter sebagai *input*.

a. Data Cleaning

Pada tahap *cleaning*, dilakukan pengecekan *Missing Value* dan dilakukan penanganan dengan mengganti nilai yang missing (tidak diketahui) sesuai data atributnya masing-masing. Dapat diketahui bahwa data yang mengalami *Missing Value* yaitu data curah hujan sebanyak 30 *record* dan data kelembapan sebanyak

30 *record* sehingga total *Missing Value* sebanyak 60 *record* dari Tabel 2 dapat dilihat beberapa data yang terdapat *Missing Value*. Data curah hujan dan kelembapan yang tidak diketahui diganti dengan nilai rata-rata dari kedua data tersebut. Proses mengganti *Missing Value* dapat dilihat pada Tabel 3 yang dilakukan dengan proses manual dengan bantuan tools Microsoft Excel.

Tabel 2. Data sebelum proses Missing Value

57.79	73.86	58.77	30.76	32.47
178	227.5	181	94.75	100
60.53	60.53	60.53	24.41	24.41
186.42	186.42	186.42	75.17	75.17
49.35	49.35	49.35	21.45	21.45
152	152	152	66.08	66.08
42.67	42.67	20.97	4.19	10.42
131.42	131.42	64.58	12.92	32.08
...
...
...

Tabel 3. Data setelah proses Missing Value

57.79	73.86	58.77	30.76	32.47
178	227.5	181	94.75	100
60.53	60.53	60.53	24.41	24.41
186.42	186.42	186.42	75.17	75.17
49.35	49.35	49.35	21.45	21.45
152	152	152	66.08	66.08
42.67	42.67	20.97	4.19	10.42
131.42	131.42	64.58	12.92	32.08
52.584	56.602	47.403	20.205	22.186
161.960	174.335	146.000	62.230	68.333

Setelah melewati proses terhadap *Missing Value*, kemudian akan dilakukan proses terhadap data outlier. Mendeteksi data outlier yaitu dengan cara mengkonversi nilai data dengan cara *standardized* (*Z-score*). Data dikatakan outlier apabila nilai *Z* yang didapat lebih besar dari 2,5 atau lebih kecil dari -2,5 ($-2,5 \leq Z \leq +2,5$). Dapat dilihat pada Tabel 4 merupakan beberapa hasil data outlier.

Tabel 4. Data Outlier

Standardized	Outlier
2.652148	Outlier

2.097741	Tidak
0.353411	Tidak
3.451791	Outlier
0.522309	Tidak
0.37822	Tidak
3.484235	Outlier
4.156966	Outlier
0.417344	Tidak
0.089089	Tidak
...	...

Terdapat 23 data yang dianggap memiliki nilai yang dianggap mengganggu. 23 data tersebut terdapat pada data kepadatan penduduk. Kemudian dilakukan penangan dengan data tersebut dengan cara penghapusan data. Berikut pada Tabel 5 merupakan hasil antara data awal dengan data setelah melewati proses data *cleaning*.

Tabel 5. Jumlah Data

Data awal	Data Bersih
150	127

b. Data Selection

Selanjutnya pada tahap Selection, dilakukan pemilihan data untuk selanjutnya dijadikan atribut yang sesuai dengan kebutuhan pada penelitian. Adapun atribut yang digunakan yaitu data curah hujan, kepadatan penduduk, kelembapan, mobilitas penduduk dan kasus DBD. Berikut pada Tabel 6 merupakan hasil dari data selection.

Tabel 6. Data Selection

Kepadatan penduduk	Mobilitas penduduk	Kelem bapan	Curah Hujan	Kasus DBD
403	0.97	53.33	164.25	Tinggi
427	0.98	53.33	164.08	Tinggi
385	0.98	53.27	157.42	Sedang
3404	0.98	53.27	146.42	Rendah
1793	2.03	58.77	233.33	Rendah
458	2.90	14.61	97.33	Tinggi
1473	1.83	14.61	78.25	Sedang

568	3.74	55.14	164.08	Sedang
...

c. Data Integration

Setelah melewati tahap data *cleaning* dan selection, atribut yang dipilih dari tahun 2016-2020 digabungkan untuk memudahkan proses mengolah data. Untuk hasil penggabungan dari data tersebut dapat dilihat pada Tabel 7.

Tabel 7. Design Interface

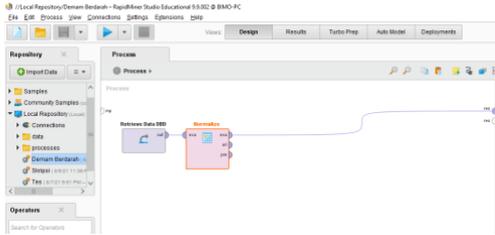
Kepadatan penduduk	Mobilitas penduduk	Kelem bapan	Curah Hujan	Kasus DBD
403	0.97	53.33	164.25	Tinggi
385	0.98	53.27	157.42	Sedang
...
1792	0.94	53.27	146.42	Rendah
838	0.84	60.53	95.75	Tinggi
...
2057	3.69	10.42	4.25	Rendah
619	2.15	54.59	118.69	Sedang
...

d. Data Transformation

pada tahapan ini dilakukan transformasi 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:

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

Dimana Z adalah variable kompleks. Normalisasi data dengan Transformasi-Z kali ini dilakukan menggunakan bantuan tools *Rapidminer*. Berikut pada Gambar 3 merupakan proses normalisasi data.



Gambar 3. Proses Normalisasi Data

Setelah melewati proses normalisasi berikut pada Tabel 8 merupakan hasil dari data transformasi yang sudah diproses pada Gambar 3.

Tabel 8. Hasil Data Transformasi

Kepadatan penduduk	Mobilitas penduduk	Kelem bapan	Curah Hujan	Kasus DBD
-1.018	-0.068	0.737	0.718	Tinggi
-1.049	-0.056	0.734	0.715	Sedang
...
0.112	-0.056	-1.215	-1.203	Rendah
-0.478	-0.056	-1.124	-1.114	Tinggi
...
-0.497	-0.231	-1.221	-1.209	Rendah
-0.922	-0.231	-0.046	-0.052	Sedang
...

4. Modelling

Pada 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 pada Tabel 9 merupakan hasil pembagian terhadap dataset.

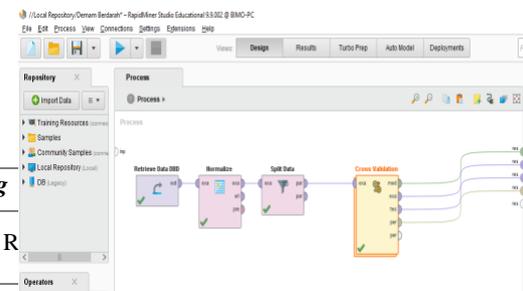
Tabel 9. Proporsi Pembagian Dataset

Dataset	Data Training			Data Testing		R
	Tinggi	Sedang	Rendah	Tinggi	Sedang	
127	66	18	18	12	5	8

Berdasarkan tabel hasil pembagian dataset diatas, diketahui bahwa data training berjumlah 102 *record* dengan class tinggi sebanyak 66 *record*, class sedang 18 *record* dan class rendah sebanyak 18 *record*. Sedangkan data testing berjumlah 25 dengan class tinggi sebanyak 12 *record*, class sedang 5 *record* dan class rendah sebanyak 8 *record*. Sehingga total jumlah class tinggi sebanyak 78 *record*, class sedang 23 *record* dan class rendah sebanyak 26 *record*.

a. 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 4. Proses Pengujian Data

Berdasarkan gambar diatas terdapat operator Cross Validation yang di dalamnya terdapat subproses. Berikut pada Gambar 5 merupakan rangkaian subproses pada cross validation. Pada subproses training terdapat operator KNN dan pada subproses testing terdapat operator apply model dan *performance*.



Gambar 5. Subproses Operator Cross Validation

Terdapat beberapa operator yang digunakan dalam proses pengujian data sebagai berikut:

- Retrieve data DBD: Operator ini dapat mengakses informasi yang disimpan di Repositori dan memuatnya ke dalam Proses.
- Normalize: Operator ini menormalisasikan nilai Atribut yang dipilih.
- Split Data: Operator ini menghasilkan pembagian data yang diinginkan dari Dataset yang diberikan. Dataset dipartisi menjadi data training dan data testing sesuai dengan ukuran relatif yang ditentukan.
- Cross Validation: Operator untuk evaluasi model.
- KNN: Operator ini menghasilkan model *K-Nearest Neighbor* yang digunakan untuk klasifikasi data.
- Apply Model: Operator ini menerapkan model pada ExampleSet.
- Performance: Operator ini digunakan untuk evaluasi kinerja statistik tugas klasifikasi. Operator ini memberikan daftar nilai kriteria

kinerja tugas klasifikasi, pada penelitian ini dilihat akurasi.

5. Evaluation

Tahap evaluasi dilakukan untuk melihat *performance* kinerja klasifikasi yang telah dilakukan pada pemodelan sebelumnya menggunakan confusion matrix untuk melihat nilai *accuracy*, recall dan precision. Berikut pada Tabel 10– Tabel 14 merupakan hasil confusion matrix dari Algoritma KNN dengan nilai K=3, K=5, K=7, K=9, dan K=11.

Tabel 10. Hasil Confusion Matrix K=3

	true Tinggi	true Sedan g	true Renda h	class precision
pred. Tinggi	58	12	2	80.56%
pred. Sedan g	5	6	1	50.00%
pred. Renda h	3	0	15	83.33%
class recall	87.88 %	33.33 %	83.33%	

Tabel 11. Hasil Confusion Matrix K=5

	true Tinggi	true Sedan g	true Renda h	class precision
pred. Tinggi	61	10	3	82.43%
pred. Sedan g	4	8	1	61.54%
pred. Renda h	1	0	14	93.33%
class recall	92.42 %	44.44 %	77.78%	

Tabel 12. Hasil Confusion Matrix K=7

	true Tinggi	true Sedan g	true Renda h	class precision
pred. Tinggi	58	10	2	82.86%
pred. Sedan g	7	8	1	50.00%
pred. Renda h	1	0	15	93.75%
class recall	87.88 %	44.44 %	83.33%	

Tabel 13. Hasil Confusion Matrix K=9

	true Tinggi	true Sedang	true Rendah	class precision
pred. Tinggi	57	10	2	82.61%
pred. Sedang	8	8	0	50.00%
pred. Rendah	1	0	16	94.12%
class recall	86.36%	44.44%	88.89%	

Tabel 14. Hasil Confusion Matrix K=11

	true Tinggi	true Sedang	true Rendah	class precision
pred. Tinggi	58	11	2	81.69%
pred. Sedang	6	7	0	53.85%
pred. Rendah	2	0	16	88.89%
class recall	87.88%	38.89%	88.89%	

Berdasarkan tabel confusion matrix yang di dapat. Dapat dilihat nilai class recall dan class precision. Berikut pada Tabel 15 merupakan nilai *accuracy* dari pengujian data.

Tabel 15. Hasil Pengujian Data

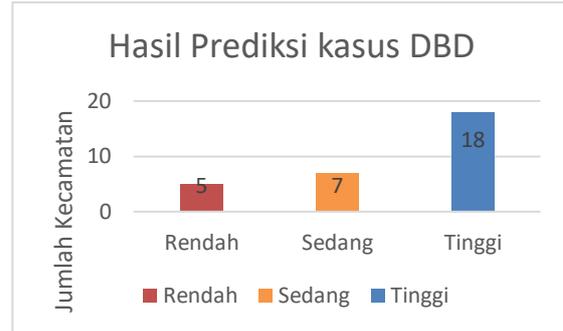
KNN	Nilai K				
	K=3	K=5	K=7	K=9	K=11
Accuracy	79.27%	81.27%	79.56%	79.45%	79.36%

Berdasarkan hasil *accuracy* diatas dapat dilihat nilai *accuracy* K=5 merupakan *accuracy* tertinggi dengan menggunakan Algoritma KNN dan *10-Fold Cross Validation* mendapatkan nilai *accuracy* sebesar 81.27%.

6. Deployment

Deployment merupakan tahap terakhir dimana hasil penelitian dipresentasikan dalam bentuk yang mudah dipahami mengenai pengetahuan yang diperoleh dalam proses *Data Mining*. Berdasarkan penelitian yang dilakukan, telah telah menghasilkan suatu informasi dan pengetahuan baru

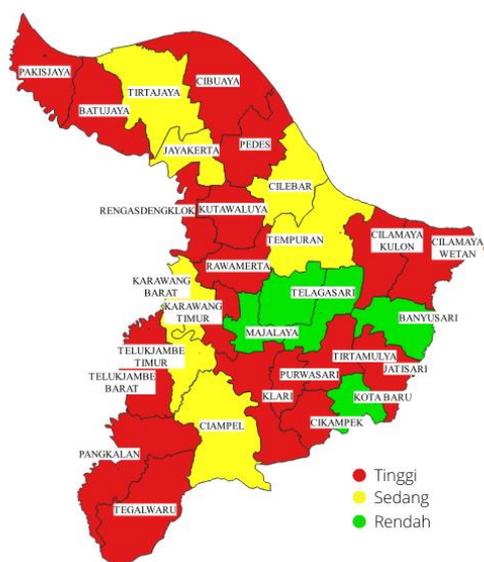
dalam proses *Data Mining* untuk klasifikasi kasus Demam Berdarah di Kabupaten Karawang. Algoritma *K-Nearest Neighbor* digunakan untuk mencari nilai akurasi yang menghasilkan prediksi kasus Demam Berdarah pada setiap kecamatan di Kabupaten Karawang. Berikut pada Gambar 6 merupakan diagram prediksi yang telah dihasilkan.



Gambar 6. Diagram Hasil Prediksi

Berdasarkan hasil prediksi tersebut, terdapat 18 Kecamatan dengan kasus demam berdarah yang terjadi relatif tinggi, 7 Kecamatan dengan kasus demam berdarah yang terjadi relatif sedang, dan 5 Kecamatan dengan kasus demam berdarah yang terjadi relatif rendah. Hal tersebut menandakan tingginya kasus Demam Berdarah di Kabupaten Karawang khususnya di 18 Kecamatan tertinggi. Berikut pada Tabel 4.23 merupakan hasil prediksi berdasarkan tiap Kecamatan.

Kemudian untuk lebih jelasnya dilakukan pemetaan dengan tools QGIS dari hasil prediksi yang sudah didapat. Berikut pada Gambar 7 merupakan hasil pemetaan dengan menggunakan tools QGIS.



Gambar 7. Hasil Pemetaan dengan GIS

SIMPULAN

Pengujian ini melakukan klasifikasi mengenai jumlah kasus demam berdarah di Kabupaten Karawang berdasarkan data pada tahun 2016 – 2020 dengan menggunakan Algoritma K-Nearest Neighbor dan metode *10-Fold Cross Validation* yang menghasilkan *performance* terbaik menggunakan nilai $K=5$ dan menghasilkan nilai *accuracy* sebesar 81.27%. Secara keseluruhan, penerapan Algoritma *K-Nearest Neighbor* dapat memberikan *accuracy* dan prediksi yang cukup baik. Adapun wilayah yang diprediksi terjadi penyebaran kasus demam berdarah yaitu 18 Kecamatan yang berpotensi tinggi, 7 Kecamatan berpotensi sedang, dan 5 Kecamatan berpotensi rendah. Hasil dari prediksi juga dapat dilihat dari pemetaan yang sudah dibuat.

DAFTAR PUSTAKA

Admojo, F. T. (2020). Klasifikasi Aroma Alkohol Menggunakan Metode Knn. *Indonesian Journal of Data and Science*, 1(2), 34–38.

Dr.J.Arunadevi, S.Ramya, & Raja, M. R. (2018). A study of classification algorithms using *Rapidminer*.

International Journal of Pure and Applied Mathematics, Volume 119(12), 15977–15988.

- Hertina, H., Nurwahid, M., Haswir, H., Sayuti, H., Darwis, A., Rahman, M., ... & Hamzah, M. L. (2021). Data mining applied about polygamy using sentiment analysis on Twitters in Indonesian perception. *Bulletin of Electrical Engineering and Informatics*, 10(4), 2231-2236.
- Kusuma, A. P., & Sukendra, D. M. (2016). Analisis Spasial Kejadian Demam Berdarah *Dengue* Berdasarkan Kepadatan Penduduk. *Unnes Journal of Public Health*, 5(1), 48. <https://doi.org/10.15294/ujph.v5i1.9703>
- Nur Itsna, I. (2020). Peningkatan Pengetahuan Masyarakat Dalam Menanggulangi Penyakit Demam Berdarah *Dengue* (DBD) di Desa Karangmalang Kedungbanteng. *JPKMI (Jurnal Pengabdian Kepada Masyarakat Indonesia)*, 1(1), 35–41. <https://doi.org/10.36596/jpkmi.v1i1.35>
- Ong, J., Liu, X., Rajarethinam, J., Kok, S. Y., Liang, S., Tang, C. S., ... Yap, G. (2018). Mapping *Dengue* risk in Singapore using Random Forest. *PLoS Neglected Tropical Diseases*, 12(6), 1–12. <https://doi.org/10.1371/journal.pntd.0006587>
- Rahim, N. F., Taib, S. M., & Abidin, A. I. Z. (2017). *Dengue* fatality prediction using *Data Mining*. *Journal of Fundamental and Applied Sciences*, 9(6S), 671-683.
- Shamrat, M. F. M. J., Chakraborty, S., Imran, M. M., Muna, J. N., Billah, M. M., Das, P., & Rahman, O. M. (2021). Sentiment analysis on twitter tweets about COVID-19

- vaccines using NLP and supervised KNN classification algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(1), 463-470.
- Sanjudevi, R., & Savitha, D. (2019). Dengue Fever Prediction using Datamining Classification Technique. *International Journal of Recent Technology and Engineering*, 8(4), 8685–8688. <https://doi.org/10.35940/ijrte.d8810.118419>.
- Sintia, S., Defit, S., & Nurcahyo, G. W. (2021). Product Codification Accuracy With Cosine Similarity And Weighted Term Frequency And Inverse Document Frequency (TF-IDF) . *Journal of Applied Engineering and Technological Science (JAETS)*, 2(2), 62–69. <https://doi.org/10.37385/jaets.v2i2.210>
- Sundari, B., & Krishnamoorthy, M. (2019). Factors to Predict Dengue Fever using Data Mining Techniques: A Review. *International Journal of Scientific Research and Engineering Development*, 2(4), 154–160. Retrieved from www.ijrsred.com
- Shokrzade, A., Ramezani, M., Tab, F. A., & Mohammad, M. A. (2021). A novel extreme learning machine based kNN classification method for dealing with big data. *Expert Systems with Applications*, 115293.