BAB III

ANALISA DAN PERANCANGAN

3.1 Analisa Sistem

Analisa dilakukan untuk memahami pola penyebaran kasus Demam Berdarah (DBD) yang tercermin pada data set dan bagaimana pola data tersebut dapat digunakan untuk membangun model klasifikasi berbasis algoritma *Naïve Bayes* dan *K-Nearest Neigbor (KNN)*. Pendektakan pada penelitian ini dimulai dengan analisis dataset yang mencakup informasi mengenai factor risiko dan pola penyearan penyakit DBD seperti, kepadatan penduduk, curah hujan dan kelembaban udara, Data ini kemudian diproses melalui tahap preprocessing untuk menghasilkan serta meningkatkan data yang berkualitas. Data yang sudah siap kemudian digunakan untuk membangun model klasifikasi berbasis dua algoritma utama diantaranya adalah Naïve Bayes algoritma ini dipilih karena kemampuanya dalam mengolah data dengan pendekatan probabilitas, sementara algoritma *KNN* dipilih karena memiliki kemampuan dalam fleksibilitas dalam menangani data numeric dan pila yang tidak berupa linear.

Setiap algoritma di uji dengan menggunakan 2 (dua) data set yang telah dibagi menjadi data latih (80%) dan data uji (20%) dengan evaluasi menggunakan metric seperti akurasi, recall, precision dan F1-Score.

$$\textbf{\textit{Precision}} = \frac{TP}{TP+FP} \times 100\%$$

$$\textbf{\textit{Recall}} = \frac{TP}{TP+FN} \times 100\%$$

$$\textbf{\textit{Akurasi}} = \frac{TP+TN}{TP+TN+FP-FN} \times 100\%$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data factor-faktor yang mempengaruhi penyebaran penyakit Demam Berdarah Dengue (DBD) di Kecamatan Aek Kuo, adapun jenis data yang digunakan adalah Data Jumlah Kasus DBD di Kecamatan Aek Kuo, Data Kepadatan Penduduk, Data Curah Hujan dan Data Kelembaban Udara di Kecamatan Aek Kuo. Data-data ini merupakan data sekunder, karena diperoleh dari Dinas Kesehatan Kecamatan Aek Kuo dan Badan Pusat Statistika Labuhanbatu Utara. Daerah atau desa yang berada di kecamatan Aek Kuo dinataranya adalah sebagai berikut:

Tabel 3.1 Desa di Kec. Aek Kuo

NO	Nama Desa
1.	Padang Maninjau
2.	Panigoran
3.	Sidomulyo
4.	Karang Anyar
5.	Padang Halaban
6.	Aek Korsik
7.	Purworejo
8.	Bandar Selamat

Data yang digunakan merupakan data yang diperoleh selama 1 tahun terakhir (2023) dengan 4 atribut penunjang keputusan pada kelas dan satu label atau kelas. Langkah pertama yang dilakukan dalam proses *Knoeledge Discovery in Database* (*KDD*) adalah seleksi data. Pada tahap ini, langkah yang diambil melibatkan pemilihan data yang menggunakan operator excel. Dibawah ini merupakan data yang telah dikumpulkan oleh peneliti.

Tabel 3.2 Hasil pengumpulan data

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
1	Padang Maninjau/Januari	4,702	112	95	Terinfeksi
2	Padang Maninjau/Februari	4,702	280	91	Tidak Terinfeksi
3	Padang Maninjau/Maret	4,702	142	95	Tidak Terinfeksi
4	Padang Maninjau/April	4,702	101	92	Tidak Terinfeksi
5	Padang Maninjau/Mei	4,702	258	92	Terinfeksi
6	Padang Maninjau/Juni	4,702	140	92	Tidak Terinfeksi
	⋮	⋮	⋮	⋮	⋮
95	Bandar Selamat/November	6,423	194	76	Tidak Terinfeksi
96	Bandar Selamat/Desember	6,423	333	73	Tidak Terinfeksi

Pada penelitian ini atribut numeric seperti kelembaban udara, curah hujan dan kepadatan penduduk memberikan informasi kualitatif sedangkan atribut daerah/bulan digunakan untuk menggambarkan desa dan bulan yang pada kejadian DBD, hal ini memberikan informasi kualitatif. Variabel-variabel ini telah diproses untuk memastikan data siap digunakan dalam algoritma klasifikasi, memastikan setiap variabel relevan terhadap tujuan penelitian.

Pada variabel kepadatan penduduk, curah hujan, kelembaban udara terbagi menjadi beberapa kategori diantaranya adalah seperti tabel berikut:

Tabel 3.2 Tabel Kategori setiap variabel

Variabel	Deskrisi	Kategori
Kepadatan Penduduk	1. 1-50 jiwa/km^2	Tidak Padat
	2. 51-250 jiwa/km^2	Kurang Padat
	3. 251-400 jiwa/km^2	Cukup Padat
	4. Lebih bersar 401 jiwa/km^2	Sangat Padat
Curah Hujan	0-100 mm	Rendah
	100-300 mm	Menengah
	300-500 mm	Tinggi
	>500 mm	Sangat Tinggi
Kelembaban Udara	0-25%	Sangat Rendah
	25-30%	Rendah
	30-60%	Normal
	60-70% >70%	Tinggi Sangat Tinggi

Tabel dibawah merupakan data yang memuat 96 data sampel dari dataset penelitian. Pada data di atas mencakup variabel variabel yang cukup relevan untuk memprediksi kasus penyebaran *Demam Berdarah Dengue* (*DBD*) seperti daerah/bulan yang memiliki data perdesa atau daerah yang memuat data setiap bulannya, Kepadatan penduduk, curah hujan, dan kelembaban udara yang rata-rata memiliki kategori yang tinggi. Data ini mencerminkan kompleksitas hubungan antar variabel dalam mempengaruhi penyebaran kasus demam berdarah. Data yang terstruktur ini mendukung penggunaan algoritma memprediksi seperti naïve bayes dan knn dalam membangun sebuah klasifikasi.

Tabel 3.4 Data Daerah terinfeksi DBD

NO	Daerah/Bulan	KP	CH	KU	Terinfeksi DBD
1	Padang Maninjau/Januari	Sangat Padat	Menengah	Terlalu Tinggi	Terinfeksi
2	Padang Maninjau/Februari	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
3	Padang Maninjau/ Maret	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
4	Padang Maninjau/April	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
5	Padang Maninjau/ Mei	Sangat Padat	Menengah	Terlalu Tinggi	Terinfeksi
6	Padang Maninjau/ Juni	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
7	Padang Maninjau/ Juli	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
8	Padang Maninjau/ Agustus	Sangat Padat	Sangat Tinggi	Terlalu Tinggi	Tidak Terinfeksi
9	Padang Maninjau/ September	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
10	Padang Maninjau/ Oktober	Sangat Padat	Menengah	Terlalu Tinggi	Tidak Terinfeksi
⋮	⋮	⋮	⋮	⋮	⋮
95	Bandar Selamat/November	Sangat Tinggi	Menengah	Terlalu Tinggi	Terinfeksi
96	Bandar Selamat/Desember	Sangat Tinggi	Tinggi	Terlalu Tinggi	Tidak Terinfeksi

Ket: KP(Kepadatan Penduduk), CH(Curah Hujan), KU(Kelembaban Udara)

3.3 Pra-pemprosesan Data

Tahap pra-pemprosesan data merupakan langkah awal yang sangat penting dalam menganalisis data, hal ini guna untuk mempersiapkan data agar dapt digunakan secara optimal dalam proses prediksi kasus *Demam Berdarah Dengue* (*DBD*). Dalam tahap *Knowlodge Discovery in Database* (*KDD*) yang dilakukan adalah transformasi. Transformasi data merupakan proses menggubah data mentah menjadi format yang lebih sesuai untuk dilakukan analisis lebih lanjut.

Tabel 3.3.1 Pembagian Kategori Transformasi Data

Variabel	Sebelum Transformasi	Sesudah Transformasi
Kepadatan Penduduk	Tidak Padat	1
	Kurang Padat	2
	Cukup Padat	3
	Sangat Padat	4
Curah Hujan	Rendah	1
	Menengah	2
	Tinggi	3
	Sangat Tinggi	4
	Sangat Rendah	5
Kelembaban Udara	Rendah	1
	Normal	2
	Tinggi	3
	Sangat Tinggi	4

Transformasi data ini mempermudah analisis yang akan di lakukan khususnya dalam penerapan algoritma *Naïve Bayes,* dengan memastikan variabel yang digunakan didalam model sudah diubah kedalam bentuk yang lebih sederhana.

Tabel 3.3.2 Nilai Hasil dari Transformasi Data

NO	Daerah/Bulan	KP	CH	KU	Status DBD
1	Padang Maninjau/Januari	4	2	5	Terinfeksi
2	Padang Maninjau/	4	2	5	Tidak Terinfeksi
3	Februari	4	2	5	Tidak Terinfeksi
4	Padang Maninjau/Maret	4	2	5	Tidak Terinfeksi
5	Padang Maninjau/April	4	2	5	Terinfeksi
6	Padang Maninjau/Mei	4	2	5	Tidak Terinfeksi
7	Padang Maninjau/Juni	4	2	5	Tidak Terinfeksi

8	Padang Maninjau/Juli	4	3	5	Tidak Terinfeksi
9	Padang Maninjau/Agustus	4	2	5	Tidak Terinfeksi
10	Padang Maninjau/Januari	4	2	5	Tidak Terinfeksi

Ket: KP(Kepadatan Penduduk), CH(Curah Hujan), KU(Kelembaban Udara)

3.4 Pengujian Metode

3.4.1 Metode Naïve Bayes

Naïve Bayes merupakan proses pengklasifikasian dengan menggunakan metode probabilitas dan statistic yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa lalu atau dimasa sebelumnya.[19] Persamaan dari *Teorema Bayes* diantaranya adalah:

$$P(C \mid F) = \frac{P(C) \times P(F \mid C)}{P(C)}$$

Keterangan:

F : Data dengan kelas yang belum diketahui

C : Hipotesis data merupakan suatu kelas spesifik

P(*C* | *F*) : Probabilitas hipotesis dengan syarat *F* (probabilitas posterior)

P(*C*) : Prtobabilitas hipotesis C (probabilitas prior)

P (*F* | *C*) : Probabilitas hipotesis *F* dengan syrat *C*

P(*F*) : Probabilitas hipotesis *F*

Sebelum melakukan pengolaan data, data set dibagi menjadi 2 bagian diantarany data training dan data testing. Dengan ketentuan data training berjumlah 80% dari jumlah seluruh data set, data training merupakan data yang digunakan sebagai data latih pada daset. Sedangkan data testing merupakan data

yang digunakan sebagai data uji yang berjumlah 20% dari jumlah seluruh data set yang ada.

Tabel 3.4 Data Training

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Status DBD
1	Padang Maninjau/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
2	Padang Maninjau/ Februari	Sangat Padat	Menengah	Tinggi	Tidak Terinfeksi
3	Padang Maninjau/Maret	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
4	Padang Maninjau/April	Sangat Padat	Menengah	Normal	Tidak Terinfeksi
5	Padang Maninjau/Mei	Sangat Padat	Menengah	Tinggi	Terinfeksi
6	Padang Maninjau/Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
7	Padang Maninjau/Juli	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
8	Padang Maninjau/Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
9	Padang Maninjau/ September	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
10	Padang Maninjau/ Oktober	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
11	Padang Maninjau/ November	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
12	Padang Maninjau/ Desember	Sangat Padat	Tinggi	Sangat Tinggi	Terinfeksi
13	Panigoran/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
14	Panigoran/Februari	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi

15	Panigoran/Maret	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
16	Panigoran/April	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
17	Panigoran/Mei	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
18	Panigoran/Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
19	Panigoran/Juli	Sangat Padat	Menengah	Tinggi	Tidak Terinfeksi
20	Panigoran/Agustus	Sangat Padat	Tinggi	Tinggi	Tidak Terinfeksi
21	Panigoran/September	Sangat Padat	Menengah	Tinggi	Terinfeksi
22	Panigoran/Oktober	Sangat Padat	Menengah	Tinggi	Terinfeksi
23	Panigoran/November	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
24	Panigoran/Desember	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
25	Sidomulyo/Januari	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
26	Sidomulyo/Februari	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
27	Sidomulyo/Maret	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
28	Sidomulyo/April	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
29	Sidomulyo/Mei	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
30	Sidomulyo/Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
31	Sidomulyo/Juli	Sangat	Menengah	Sangat Tinggi	Tidak Terinfeksi

		Padat			
32	Sidomulyo/Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi

Data diatas merupakan data training atau data latih yang digunakan untuk membantu proses klasifikasi dengan menggunakan metode naïve bayes. Data diatas nantinya juga akan dibagi menjadi beberapa bagian per variabel guna untuk mencari probabilitas pada setiap variabel.

Tabel 3.5 Data Testing

NO	**Daerah/Bulan**	**Kepadatan Penduduk**	**Curah Hujan**	**Kelembaban Udara**
1	Panigoran/Oktober	Sangat Padat	Menengah	Tinggi
2	Panigoran/November	Sangat Padat	Menengah	Sangat Tinggi
3	Padang Maninjau/Maret	Sangat Padat	Menengah	Sangat Tinggi
4	Padang Maninjau/April	Sangat Padat	Menengah	Normal
5	Panigoran/Januari	Sangat Padat	Sangat Tinggi	Sangat Tinggi
6	Sidomulyo/Juni	Sangat Padat	Menengah	Sangat Tinggi
7	Sidomulyo/Juli	Sangat Padat	Menengah	Sangat Tinggi
8	Sidomulyo/Agustus	Sangat Padat	Tinggi	Sangat Tinggi

Pada tabel diatas merupakan data uji yang akan digunakan untuk menguji kinerja model setelah dilatih. Data uji merupakan data yang labelnya belum diketahui oleh model algoritma. Model algoritma akan memprediksi akan memprediksi berdasarkan nilai yang telah dilatih. Hasil prediksi tersebut nantinya

akan dibandingkan dengan label yang sebenarnya untuk mengevaluasi akurasi model. Data tersebut akan dihitung probabilitasnya untuk menentukan daerah wilayah mana yang paling banyak terinfeksi demamberdarah di Kec. Aek Kuo.

Tabel 3.6 Variabel Daerah/Bulan

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
	Padang Maninjau/ Januari	1	0	1/7=0.142857143	0/25=0
	Padang Maninjau /Februari	0	1	0/7=0	0/25=0
	Padang Maninjau /Maret	0	1	0/7=0	0.04
	Padang Maninjau /April	0	1	0/7=0	1/25=0.04
	Padang Maninjau /Mei	1	0	1/7=0.142857143	0/25=0
	Padang Maninjau /Juni	0	1	0/7=0	1/25=0.04
	Padang Maninjau /Juli	0	1	0/7=0	1/25=0.04
	Padang Maninjau /Agustus	0	1	0/7=0	1/25=0.04
	Padang Maninjau/ September	0	1	0/7=0	1/25=0.04
	Padang Maninjau / Oktober	0	1	0/7=0	1/25=0.04
	Padang Maninjau/ November	0	1	0/7=0	1/25=0.04
	Padang Maninjau/ Desember	1	0	1/7=0.142857143	0/25=0

Daerah/ Bulan	Panigoran /Januari	1	0	1/7=0.142857143	0/25=0
	Panigoran /Februari	0	1	0/7=0	1/25=0.04
	Panigoran /Maret	0	1	0/7=0	1/25=0.04
	Panigoran /April	0	1	0/7=0	1/25=0.04
	Panigoran /Mei	0	1	0/7=0	1/25=0.04
	Panigoran /Juni	0	1	0/7=0	1/25=0.04
	Panigoran /Juli	0	1	0/7=0	1/25=0.04
	Panigoran /Agustus	0	1	0/7=0	1/25=0.04
	Panigoran /September	1	0	1/7=0.142857143	0/25=0
	Panigoran /Oktober	1	0	1/7=0.142857143	0/25=0
	Panigoran /November	1	0	1/7=0.142857143	0/25=0
	Panigoran /Desember	0	1	0/7=0	1/25=0.04
	Sidomulyo /Januari	0	1	0/7=0	1/25=0.04
	Sidomulyo /Februari	0	1	0/7=0	1/25=0.04
	Sidomulyo /Maret	0	1	0/7=0	1/25=0.04
	Sidomulyo /April	0	1	0/7=0	1/25=0.04
	Sidomulyo /Mei	0	1	0/7=0	1/25=0.04
	Sidomulyo /Juni	0	1	0/7=0	1/25=0.04
	Sidomulyo /Juli	0	1	0/7=0	1/25=0.04
	Sidomulyo /Agustus	0	1	0/7=0	1/25=0.04
	TOTAL	7	25	100%	100%

Tabel 3.7 Variabel Kepadatan Penduduk

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
Kepadatan Penduduk	Tidak Padat	0	0	0/7=0	0/25=0
	Kurang Padat	0	0	0/7=0	0/25=0
	Cukup Padat	0	0	0/7=0	0/25=0
	Sangat Padat	7	25	7/7=1	25/25=1
	Total	7	25	100%	100%

Tabel 3.8 Variabel Curah Hujan

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
Curah Hujan	Rendah	0	0	0/7=0	0/25=0
	Menengah	6	21	6/7=0.857142857	6/21=0.84
	Tinggi	1	4	1/7=0.142857143	4/250.16
	Sangat Tinggi	0	0	0/7=0	0/250
	TOTAL	7	25	100%	100%

Tabel 3.9 Tabel Kelembaban Udara

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
Kelembaban Udara	Sangat Rendah	0	0	0/7=0	0/25=0
	Rendah	0	0	0/7=0	0/25=0
	Normal	0	1	0/7=0	1/25=0.04
	Tinggi	3	3	0/7=0.428571429	3/25=0.12
	Sangat Tinggi	4	21	0/7=0.571428571	21/25=0.84
	Total	7	25	100%	100%

Tabel 3.10 Variabel Stasus DBD

Variabel		P(Terinfeksi/Tidak Terinfeksi)
Satus DBD	**Terinfeksi**	0.21875
	Tidak Terinfeksi	0.78125
Total		100%

3.4.1.1 Pengelolaan Data Menggunakan Naïve Bayes

Setelah data selesai dikelompokkan berdasarkan atributnya, maka dari itu data bisa diolah dengan menggukan metode *Naïve Bayes*. Dibawah ini merupakan pengolaan data dengan menggunakan data pertama diantaranya sebagai berikut:

$P_{(Terinfeksi)}$ =P(Daerah/Bulan|Panigoran/Agustus) × P(Kepadatan Penduduk|Sangat Padat) × P(Curah Hujan|Menengah) × P(Kelembaban Udara|Tinggi) × P(Status DBD|Terinfeksi)

= 0,1428 * 1 * 0,857 *0,428 * 0,218

=0,0113

$P_{(Tidak\ Terinfeksi)}$ =P(Daerah/Bulan|Panigoran/Agustus) × P(Kepadatan Penduduk|Sangat Padat) × P(Curah Hujan|Menengah) × P(Kelembaban Udara|Tinggi) × P(Status DBD|Tidak Terinfeksi)

= 0*1*0,84*0,12*0,781

= 0

Pada perhitungan data diatas di di dapatkan nilai yang tertinggi dimiliki oleh status *DBD* terinfeksi, hal ini dapat disimpulkan jika pada data pertama atau ke-1 merupakan daerah atau wilayah yang terinfeksi penyebaran kasus demam berdarah *dengue* (*DBD*). Pada perhitungan selanjutnya dapat dilakukan seperti perhitungan yang ada diatas, adapun hasil yang telah didapatkan dari perhitungan yang telah dilakukan adalah sebagai berikut:

Tabel 3.11 Hasil klasifikasi Menggunakan *Naïve Bayes*

NO	**Daerah/Bulan**	**Status DBD**		**Status DBD**	**Terinfeksi**	**Tidak Terinfeksi**
1	Panigoran /Oktober	Terinfeksi		TERINFEKSI	0.01148	0
2	Panigoran /November	Terinfeksi		TERINFEKSI	0.015306	0
3	Padang Maninjau /Maret	Tidak Terinfeksi	Prediksi	TIDAK TERINFEKSI	0	0.00315
4	Padang Maninjau /April	Tidak Terinfeksi		TIDAK TERINFEKSI	0	0.000294
5	Panigoran /Januari	Terinfeksi		TIDAK TERINFEKSI	0	0
6	Sidomulyo /Juni	Tidak Terinfeksi		TIDAK TERINFEKSI	0	0.02205
7	Sidomulyo /Juli	Tidak Terinfeksi		TIDAK TERINFEKSI	0	0.02205
8	Sidomulyo /Agustus	Terinfeksi		TIDAK TERINFEKSI	0	0.0042

Data diatas merupakan hasil perhitungan manual yang telah dilakukan dengan menggunakan *Microsoft Excel* dengan mencantumkan hasil prediksi disetiap data uji. Pada daerah Panigoran dengan Kepadatan Penduduk 1,570km^2 dan Curah Hujan sebesar 194mm dan Kelembaban Udara 71% maka dinyatakan berisiko tinggi terinfeksi.

3.4.1.2 Evaluasi Model *Naïve Bayes*

Tabel 3.4.1.2 Confusion Matrix *Naïve Bayes*

	CLASS	
Predicted	Terinfeksi	Tidak Terinfeksi
Terinfeksi	2 (TP)	0 (FP)
Tidak Terinfeksi	2 (FN)	4 (TN)

Tabel diatas merupakan hasil dari uji performa dengan menggunakan metode *Naïve Bayes* dengan keterangan *True Positive* (*TP*) yaitu 2, *True Negative* (*TN*) yaitu 4, *False Negative* (*TN*) yaitu 2 dan *False Positive* (*FP*) yaitu 0.

Adapun rumus yang digunakan untuk mencari sebuah *accuracy* adalah sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

$$= \frac{2+4}{2+0+2+4}$$

$$= \frac{6}{8} = 0.75 * 100\% = 75\%$$

Adapun rumus yang digunakan untuk mencari sebuah *Precision* adalah sebagai berikut :

$$Precison = \frac{TP}{TP+FP}$$

$$= \frac{2}{2+0}$$

$$= \frac{2}{2} = 1 * 100\% = 100\%$$

Adapun rumus yang digunakan untuk mencari sebuah *Recall* adalah sebagai berikut :

$$Recall = \frac{TP}{TP+FN}$$

$$= \frac{2}{2+2}$$

$$= \frac{2}{4} = 0.5 * 100\% = 50\%$$

Adapun rumus yang digunakan untuk mencari sebuah *F1-Score* adalah sebagai berikut :

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

$$= 2 \times \frac{100\% \times 50\%}{100\%+50\%}$$

$$2 \times \frac{5000\%}{150\%} = 66{,}666 = 67\%$$

Setelah mengetahui hasil dari uji peforma di atas, maka klasifikasi daerah atau wilayah yang terinfeksi *DBD* dengan menggunakan metode *Naïve Bayes* dinyatakan baik dan akurat hal ini dikarenakan memiliki tingkat akurasi yang cukup tinggi.

3.4.2 Metode *K-nearest Neighbor*

Algoritma K-NN merupakan suatu metode klasifikasi data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. Prinsip kerja dari metode K-NN ini adalah dengan cara menentukan K objek yang berbeda pada data pelatihan dengan jarak yang mendekati atau paling dekat

dengan objek atau data test. Proses perhitungan didalam algoritma K-NN ini merupakan sebuah perhitungan *Euclidean Distance* yaitu metode pencarian antara dua titik variabel, semakin dekat dan mirip maka semakin kecil jarak antara dua titik tersebut. Perhitungan jarak dilakukan dengan rumus *Eulidean Distance* yang ditunjukkan pada persamaan berikut:

$$d(x,y) = \sqrt{\sum_{i}^{n} (x^{i}_{training} - y^{i}_{testing})^2}$$

Keterangan:

d (x,y) = jarak

$x^{i}_{training}$ = *Data testing*

$y^{i}_{testing}$ = *Data training*

i = Variabel data

n = Dimensi data

Sebelum melakukan pengolaan data, data set dibagi menjadi 2 bagian diantarany data training dan data testing. Dengan ketentuan data training berjumlah 80% dari jumlah seluruh data set, data training merupakan data yang digunakan sebagai data latih pada daset. Sedangkan data testing merupakan data yang digunakan sebagai data uji yang berjumlah 20% dari jumlah seluruh data set yang ada.

Tabel 3.12 Data Training *K-Nearest Neighbor*

Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
Padang Maninjau/Januari	4,702	112	70	Terinfeksi
Padang Maninjau/Februari	4,702	280	72	Tidak Terinfeksi

Padang Maninjau/Maret	4,702	142	80	Tidak Terinfeksi
Padang Maninjau/April	4,702	101	80	Tidak Terinfeksi
Padang Maninjau/Mei	4,702	258	72.6	Terinfeksi
Padang Maninjau/Juni	4,702	140	77	Tidak Terinfeksi
Padang Maninjau/Juli	4,702	150	85	Tidak Terinfeksi
Padang Maninjau/Agustus	4,702	471	73	Tidak Terinfeksi
Padang Maninjau/September	4,702	153	93	Tidak Terinfeksi
Padang Maninjau/Oktober	4,702	185	92	Tidak Terinfeksi
Padang Maninjau/November	4,702	194	72	Tidak Terinfeksi
Padang Maninjau/Desember	4,702	333	89	Terinfeksi
Panigoran/Januari	1,570	112	80	Terinfeksi
Panigoran/Februari	1,570	280	80.2	Tidak Terinfeksi
Panigoran/Maret	1,570	142	78.4	Tidak Terinfeksi
Panigoran/April	1,570	101	78.4	Tidak Terinfeksi
Panigoran/Mei	1,570	258	73.8	Tidak Terinfeksi
Panigoran/Juni	1,570	140	71.9	Tidak Terinfeksi
Panigoran/Juli	1,570	150	68.5	Tidak Terinfeksi
Panigoran/Agustus	1,570	471	68.6	Tidak Terinfeksi
Panigoran/September	1,570	153	68.7	Terinfeksi
Panigoran/Oktober	1,570	185	69	Terinfeksi
Panigoran/November	1,570	194	71	Terinfeksi
Panigoran/Desember	1,570	333	79.1	Tidak Terinfeksi
Sidomulyo/Januari	1,907	112	89.3	Tidak Terinfeksi
Sidomulyo/Februari	1,907	280	85.8	Tidak Terinfeksi
Sidomulyo/Mart	1,907	142	87.8	Tidak Terinfeksi
Sidomulyo/April	1,907	101	86	Tidak Terinfeksi
Sidomulyo/Mei	1,907	258	81.3	Tidak Terinfeksi
Sidomulyo/Juni	1,907	140	79.3	Tidak Terinfeksi
Sidomulyo/Juli	1,907	150	76.5	Tidak Terinfeksi
Sidomulyo/Agustus	1,907	471	76.5	Tidak Terinfeksi

Data diatas merupakan data training yang digunakan untuk melatih model. Berisi pasangan input(fitur) dan aoutput (label) yang telah diketahui. Metode *KNN* akan mempelajari distribusi Probabilitas dari data laiah lain guna untuk melakukan prediksi.

Tabel 3.13 Data Testing *K-Nearest Neighbor*

NO	**Daerah/Bulan**	**Kepadatan Penduduk**	**Curah Hujan**	**Kelembaban Udara**	**Status DBD**
1	Panigoran/Oktober	1,570	185	69	Terinfeksi
2	Panigoran/November	1,570	194	71	Terinfeksi
3	Padang Maninjau/Maret	4,702	142	95	Tidak Terinfeksi
4	Padang Maninjau/April	4,702	101	92	Tidak Terinfeksi
5	Panigoran/Januari	1,570	112	80	Terinfeksi
6	Sidomulyo/Juni	1,907	140	79.3	Tidak Terinfeksi
7	Sidomulyo/Juli	1,907	150	76.5	Tidak Terinfeksi
8	Sidomulyo/Agustus	1,907	471	76.5	Tidak Terinfeksi

Data diatas merupakan data uji yang digunakan untuk menguji kinerja model *K-NN*. Data tersebut berisi data yang lebelnya belum diketahui. *K-NN* akan memprediksi berdasarkan apa yang telah dipelajari dari data latih, dan hasil ini nantinya akan dibandingkan dengan label yang sebenarnya untuk mengevaluasi model. Selanjutnya data akan dihitung jarak encludiean untuk memprediksi wilayah atau daerah mana yang terinfeksi *Demam Berdarah Dengue* (DBD).

3.4.2.1 Pengolaan Data Menggunakan Metode K-Nearest Neighbor

Tabel 3.14 Data Baru

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara
1	Bandar Selamat/Agustus	1,305	98	60

Pada tabel diatas merupakan data baru yang nantinya akan di prediksi apakah daerah tersebut termasuk merupakan daerah atau wilayah yang terinfeksi atau tidak. Selanjutnya kita menghitung jarak terdekat sebagai berikut:

$$\sqrt{(1{,}570 - 1{,}305)^2 + (185 - 98)^2 + (69 - 60)^2}$$

$$= \sqrt{77.874}$$

$$=279.060$$

Selanjutnya lakukan perhitungan jarak seperti diatas hingga ke data yang terakhir atau yang ke 8. Kemudian urutkan dari jarak yang terkecil hingga yang terbesar seperti tabel berkut:

Tabel 3.15 Data dengan menentukan jarak terdekat

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Status DBD	Jarak Encludiean
1	Panigoran /Oktober	1,570	185	69	Terinfeksi	1.136298201
2	Panigoran /November	1,570	194	71	Terinfeksi	1.323812389
3	Padang Maninjau /Maret	4,702	142	95	Tidak Terinfeksi	4.100637856
4	Padang Maninjau /April	4,702	101	92	Tidak Terinfeksi	3.869723252
5	Panigoran /Januari	1,570	112	80	Terinfeksi	1.85134476
6	Sidomulyo	1,907	140	79.3	Tidak	1.863811831

	/Juni				Terinfeksi	
7	Sidomulyo /Juli	1,907	150	76.5	Tidak Terinfeksi	1.642843136
8	Sidomulyo /Agustus	1,907	471	76.5	Tidak Terinfeksi	3.60012346

Selanjutnya tentukan niali K, dimana nilai K yang ditentukan yaitu K=3. Maka klasifikasi diambil 3 tetangga terdekat.

Tabel 3.16 Hasil Nilai nilai *K*

NO	Daerah /Bulan	KP	CH	KU	Status DBD	Daerah/ Bulan	Encludiean
1	Padang Maninjau /Maret	4,702	142	95	Tidak Terinfeksi	Padang Maninjau /Maret	1
2	Padang Maninjau /April	4,702	101	92	Tidak Terinfeksi	Padang Maninjau /April	2
3	Sidomulyo /Agustus	1,907	471	76.5	Tidak Terinfeksi	Sidomulyo /Agustus	12

Ket: KP(Kepadatan Penduduk), CH(Curah Hujan), KU(Kelembaban Udara)

Terlihat pada tabel diatas dengan menentukan k=3 atau 3 tetangga terdekat dari data atau nilai terkecil menunjukkan bahwa klasifikasi data uji 1 sebanyak 3 *Tidak Terinfeksi* dan 0 *Terinfeksi.* Maka dari itu dapat disimpulkan bahwa data uji ke 1 tidak terinfeksi DBD. Selanjutnya perhitungan dilakukan dengan cara yang sama dengan mencari tetangga terdekat dari setiap data uji yang baru. Berikut merupakan hasil akhir klasifikasi dari data uji :

Tabel 3.17 Hasil Klasifikasi

NO	**Daerah/Bulan**	**Status DBD**		**Satus DBD**
1	Panigoran/Oktober	Terinfeksi		Tidak Terinfeksi
2	Panigoran/November	Terinfeksi		Tidak Terinfeksi
3	Padang Maninjau/Maret	Tidak Terinfeksi		Tidak Terinfeksi
4	Padang Maninjau/April	Tidak Terinfeksi	Prediksi	Tidak Terinfeksi
5	Panigoran/Januari	Terinfeksi		Terinfeksi
6	Sidomulyo/Juni	Tidak Terinfeksi		Tidak Terinfeksi
7	Sidomulyo/Juli	Tidak Terinfeksi		Terinfeksi
8	Sidomulyo/Agustus	Tidak Terinfeksi		Terinfeksi

Hasil klasifikasi dengan metode *K-Nearest Neighbor* (*KNN*) menunjukkan bahwa dari 8 data uji ada 3 wilayah yang terinfeksi penyakit DBD dan 5 wilayah yang tidak terinfeksi kasus Demam Berdarah (*DBD*). Pada daerah Panigoran Kepadatan Penduduk 1,570 km^2 dan Curah Hujan sebesar 112mm serta Kelembaban Udara 80% maka dinyatakan berisiko tinggi terinfeksi *DBD.* Sedangkan pada daerah Padang Maninjau dengan Kepadatan Penduduk 4,702 km^2 dan Curah Hujan 112mm serta Kelembaban Udara 70% maka dinyatakan tidak terinfeksi *DBD.*

3.4.2.2 Evaluasi Model (Conffusion Matrix) K-Nearest Neighbor

Tabel 3.2.1 Confusion Matrix K-Nearest Neighbor

	CLASS	
Predicted	Terinfeksi	Tidak Terinfeksi
Terinfeksi	1(TP)	2(FP)
Tidak Terinfeksi	2(FN)	3(TN)

Tabel diatas merupakan hasi dari ui peforma dengan menggunakan metode *K-Nearest Neighbor*, dimana menghasilkan *True Positive (TP) 1*, *True Negative (TN) 3*, *False Positive (FP) 2*, dan *False Negative (FN)* berjumlah 2.

Adapun rumus yang digunakan untuk mencari sebuah *accuracy* adalah sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

$$= \frac{1+3}{1+2+2+3}$$

$$= \frac{4}{8} = 0.5 * 100\% = 50\%$$

Adapun rumus yang digunakan untuk mencari sebuah *Precision* adalah sebagai berikut :

$$Precison = \frac{TP}{TP+FP}$$

$$= \frac{1}{2+2}$$

$$= \frac{1}{2} = 0{,}25 * 100\% = 25\%$$

Adapun rumus yang digunakan untuk mencari sebuah *Recall* adalah sebagai berikut :

$$Recall = \frac{TP}{TP+FN}$$

$$= \frac{1}{1+2}$$

$$=\frac{3}{1} = 0.33 * 100\% = 33\%$$

Adapun rumus yang digunakan untuk mencari sebuah *F1-Score* adalah sebagai berikut :

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$= 2 \times \frac{25\% \times 33\%}{25\% + 33\%}$$

$$2 \times \frac{825\%}{58\%} = 14 * 100\% = 14\%$$

Terlihat dari hasil perhitungan yang telah dilakukan dengan melakukan perbandingan dari kedua metode menunjukkan bahwa metode *Naïve Bayes* lebih unggul dalam prediksi penyebaran kasus *Demam Berdarah Dengue (DBD)* dengan perbandingan akurasi 75% sedangkan *K-NN* memiliki akurasi 50%. Jadi dapat disimpulkan jika algoritma *Naïve Bayes* lebih baik dalam klasifikasi kasus ini.