

BAB IV

IMPLEMENTASI DAN PEMBAHASAN

4.1 Penerapan Metode Naïve Bayes

Ada pun beberapa langkah yang dapat dilakukan dalam penerapan metode *Naïve Bayes* diantaranya adalah sebagai berikut:

4.1.1 Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data sekunder. Data sekunder sendiri merupakan sumber data yang didapatkan oleh peneliti baik melalui media perantara atau tidak secara langsung. Data sekunder yang diperoleh dari Dinas Kesehatan Kec. Aek Kuo yaitu data jumlah kasus terinfeksi *DBD* perbulan selama 1 tahun terakhir (2023) yang terdiri dari 8 desa yang ada di Kec. Aek Kuo. Sedangkan data sekunder yang diperoleh dari Badan Pusat Statistika Labuhanbatu Utara yaitu data Kepadatan Penduduk di setiap desa, Curah Hujan dan Kelmbaban Udara. Maka dapat diperoleh jumlah data yang digunakan sebanyak 96 data.

Tabel 4.1 View Data

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
1	Padang Maninjau/Januari	4,702	112	70	Terinfeksi
2	Padang Maninjau/Februari	4,702	280	72	Tidak Terinfeksi
3	Padang Maninjau/Maret	4,702	142	80	Tidak Terinfeksi
4	Padang Maninjau/April	4,702	101	80	Tidak Terinfeksi
5	Padang Maninjau/Mei	4,702	258	72.6	Terinfeksi
6	Padang Maninjau/Juni	4,702	140	77	Tidak Terinfeksi
7	Padang Maninjau/Juli	4,702	150	85	Tidak Terinfeksi

8	Padang Maninjau/Agustus	4,702	471	73	Tidak Terinfeksi
9	Padang Maninjau/September	4,702	153	93	Tidak Terinfeksi
10	Padang Maninjau/Okttober	4,702	185	92	Tidak Terinfeksi
11	Padang Maninjau/November	4,702	194	72	Tidak Terinfeksi
12	Padang Maninjau/Desember	4,702	333	89	Terinfeksi
13	Panigoran/Januari	1,570	112	80	Terinfeksi
14	Panigoran/Februari	1,570	280	80.2	Tidak Terinfeksi
15	Panigoran/Maret	1,570	142	78.4	Tidak Terinfeksi
16	Panigoran/April	1,570	101	78.4	Tidak Terinfeksi
17	Panigoran/Mei	1,570	258	73.8	Tidak Terinfeksi
18	Panigoran/Juni	1,570	140	71.9	Tidak Terinfeksi
19	Panigoran/Juli	1,570	150	68.5	Tidak Terinfeksi
20	Panigoran/Agustus	1,570	471	68.6	Tidak Terinfeksi
21	Panigoran/September	1,570	153	68.7	Terinfeksi
22	Panigoran/Okttober	1,570	185	69	Terinfeksi
23	Panigoran/November	1,570	194	71	Terinfeksi
24	Panigoran/Desember	1,570	333	79.1	Tidak Terinfeksi
25	Sidomulyo/Januari	1,907	112	89.3	Tidak Terinfeksi
26	Sidomulyo/Februari	1,907	280	85.8	Tidak Terinfeksi
27	Sidomulyo/Mart	1,907	142	87.8	Tidak Terinfeksi
28	Sidomulyo/April	1,907	101	86	Tidak Terinfeksi
29	Sidomulyo/Mei	1,907	258	81.3	Tidak Terinfeksi
30	Sidomulyo/Juni	1,907	140	79.3	Tidak Terinfeksi
31	Sidomulyo/Juli	1,907	150	76.5	Tidak Terinfeksi
32	Sidomulyo/Agustus	1,907	471	76.5	Tidak Terinfeksi
33	Sidomulyo/September	1,907	153	77.8	Terinfeksi
34	Sidomulyo/Okttober	1,907	185	75.3	Terinfeksi
35	Sidomulyo/November	1,907	194	77	Terinfeksi
36	Sidomulyo/Desember	1,907	333	92.3	Tidak Terinfeksi
37	Karang Anyar/Januari	1,036	112	65	Terinfeksi
38	Karang Anyar/Februari	1,036	280	77	Tidak Terinfeksi
39	Karang Anyar/Maret	1,036	142	77	Tidak Terinfeksi
40	Karang Anyar/April	1,036	101	65	Tidak Terinfeksi
41	Karang Anyar/Mei	1,036	258	65	Terinfeksi
42	Karang Anyar/Juni	1,036	140	72	Tidak Terinfeksi
43	Karang Anyar/Juli	1,036	150	68	Tidak Terinfeksi
44	Karang Anyar/Agustus	1,036	471	98	Tidak Terinfeksi

45	Karang Anyar/September	1,036	153	76	Terinfeksi
46	Karang Anyar/Oktober	1,036	185	70	Tidak Terinfeksi
47	Karang Anyar/November	1,036	194	67	Tidak Terinfeksi
48	Karang Anyar/Desember	1,036	333	98	Tidak Terinfeksi
49	Padang Halaban/Januari	1,500	112	82	Tidak Terinfeksi
50	Padang Halaban/Februari	1,500	280	83	Tidak Terinfeksi
51	Padang Halaban/Maret	1,500	142	78	Tidak Terinfeksi
52	Padang Halaban/April	1,500	101	78	Tidak Terinfeksi
53	Padang Halaban/Mei	1,500	258	75	Tidak Terinfeksi
54	Padang Halaban/Juni	1,500	140	74	Tidak Terinfeksi
55	Padang Halaban/Juli	1,500	150	71	Tidak Terinfeksi
56	Padang Halaban/Agustus	1,500	471	70	Tidak Terinfeksi
57	Padang Halaban/September	1,500	153	70	Terinfeksi
58	Padang Halaban/Oktober	1,500	185	73	Tidak Terinfeksi
59	Padang Halaban/November	1,500	194	75	Tidak Terinfeksi
60	Padang Halaban/Desember	1,500	333	74	Tidak Terinfeksi
61	Aek Korsik/Januari	15,976	112	70	Terinfeksi
62	Aek Korsik/Februari	15,976	280	62	Tidak Terinfeksi
63	Aek Korsik/Maret	15,976	142	87	Tidak Terinfeksi
64	Aek Korsik/April	15,976	101	80	Tidak Terinfeksi
65	Aek Korsik/Mei	15,976	258	65	Tidak Terinfeksi
66	Aek Korsik/Juni	15,976	140	59.2	Tidak Terinfeksi
67	Aek Korsik/Juli	15,976	150	78	Tidak Terinfeksi
68	Aek Korsik/Agustus	15,976	471	89	Terinfeksi
69	Aek Korsik/September	15,976	153	90	Terinfeksi
70	Aek Korsik/Oktober	15,976	185	89	Terinfeksi
71	Aek Korsik/November	15,976	194	85	Terinfeksi
72	Aek Korsik/Desember	15,976	333	92	Tidak Terinfeksi
73	Purworejo/Januari	1,570	112	81.8	Tidak Terinfeksi
74	Purworejo/Februari	1,570	280	83	Tidak Terinfeksi
75	Purworejo/Maret	1,570	142	78.3	Tidak Terinfeksi
76	Purworejo/April	1,570	101	76.1	Tidak Terinfeksi
77	Purworejo/Mei	1,570	258	76	Tidak Terinfeksi
78	Purworejo/Juni	1,570	140	79	Tidak Terinfeksi

79	Purworejo/Juli	1,570	150	72.8	Tidak Terinfeksi
80	Purworejo/Agustus	1,570	471	72.1	Tidak Terinfeksi
81	Purworejo/September	1,570	153	72.7	Terinfeksi
82	Purworejo/Oktober	1,570	185	73.5	Terinfeksi
83	Purworejo/November	1,570	194	76.3	Terinfeksi
84	Purworejo/Desember	1,570	333	79.6	Tidak Terinfeksi
85	Bandar Selamat/Januari	6,423	112	75	Tidak Terinfeksi
86	Bandar Selamat/Februari	6,423	280	81	Tidak Terinfeksi
87	Bandar Selamat/Maret	6,423	142	80	Tidak Terinfeksi
88	Bandar Selamat/April	6,423	101	75	Tidak Terinfeksi
89	Bandar Selamat/Mei	6,423	258	76	Tidak Terinfeksi
90	Bandar Selamat/Juni	6,423	140	77	Tidak Terinfeksi
91	Bandar Selamat/Juli	6,423	150	72	Tidak Terinfeksi
92	Bandar Selamat/Agustus	6,423	471	72	Tidak Terinfeksi
93	Bandar Selamat/September	6,423	153	67	Terinfeksi
94	Bandar Selamat/Oktober	6,423	185	69	Terinfeksi
95	Bandar Selamat/November	6,423	194	76	Tidak Terinfeksi
96	Bandar Selamat/Desember	6,423	333	73	Tidak Terinfeksi

Data asli yang didapatkan penulis memiliki 4 atribut diantaranya yaitu daerah/bulan, kepadatan penduduk, curah hujan, kelembaban udara.

4.1.2 Pembersihan Data

Setelah tahap pengumpulan data dan filter data maka tahap selanjutnya yaitu cleaning data agar tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses data mining. Semua data yang dibutuhkan telah melalui tahap cleaning data maka penulis mendapatkan 96 data yang akan diolah menjadi 2 bagian yaitu *data*

training dan *data testing* yang akan dimasukkan atau diolah melalui bantuan software yang digunakan pada penelitian ini.

4.1.3 Transformasi Data

Data Transformation adalah tahap mengubah data menjadi bentuk yang sesuai guna untuk diproses dalam *data mining*. Dalam penelitian ini data yang akan diproses dari *Microsoft Excel 2010* akan digunakan untuk pengolahan pada *Software RapidMiner*

Selanjutnya penulis melakukan transformasi data agar penelitian ini bisa berjalan dengan baik. Data yang akan dibagi menjadi beberapa kategori merupakan data dari atribut kepadatan penduduk, curah hujan, dan kelembaban udara berikut ini:

Tabel 4.1 **Kategori variabel kepadatan penduduk**

Variabel	Deskripsi	Kategori
Kepadatan Penduduk	1-50 jiwa/km ²	Tidak Padat
	51-250 jiwa/km ²	Kurang Padat
	251-400 jiwa/km ²	Padat
	> 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	1 – 25%	Sangat Rendah
	25 -30%	Rendah
	30 – 60 %	Normal
	>70%	Sangat Tinggi

Pada gambar dibawah ini merupakan data yang sudah dibagi berdasarkan kategorinya:

	A	B	C	D	E
1	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
2	Padang Maninjau/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
3	Padang Maninjau/Februari	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
4	Padang Maninjau/Maret	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
5	Padang Maninjau/April	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
6	Padang Maninjau/Mei	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
7	Padang Maninjau/Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
8	Padang Maninjau/Juli	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
9	Padang Maninjau/Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
10	Padang Maninjau/September	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
11	Padang Maninjau/Oktober	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
12	Padang Maninjau/November	Sangat Padat	Menengah	Tinggi	Tidak Terinfeksi
13	Padang Maninjau/Desember	Sangat Padat	Tinggi	Sangat Tinggi	Terinfeksi
14	Panigoran/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
15	Panigoran/Februari	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
16	Panigoran/Maret	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
17	Panigoran/April	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
18	Panigoran/Mei	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
19	Panigoran/Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
20	Panigoran/Juli	Sangat Padat	Menengah	Tinggi	Tidak Terinfeksi
21	Panigoran/Agustus	Sangat Padat	Tinggi	Tinggi	Tidak Terinfeksi
22	Panigoran/September	Sangat Padat	Menengah	Tinggi	Terinfeksi
23	Panigoran/Oktober	Sangat Padat	Menengah	Tinggi	Terinfeksi
24	Panigoran/November	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
25	Panigoran/Desember	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi

Gambar 4.1.2 Data kasus terinfeksi *DBD*

4.1.4 Implementasi Dengan Metode Naïve Bayes

Pada tahap ini pemodelan data, metode yang digunakan pada penelitian ini adalah probabilitas (prediksi) dengan menggunakan algoritma *Naïve Bayes* Data yang telah dikumpulkan ditransformasi akan dikelola menggunakan probabilitas dan perhitungan jarak. Metode ini dapat digunakan dalam memprediksi peluang yang akan terjadi dimasa depan berdasarkan pengalaman dimasa sebelumnya sebagai bahan perbandingan.

Data yang akan diujikan dibagi menjadi dua yaitu data *training* dan *testing* kemudian dianalisis dengan menggunakan *Software RapidMiner*. Data Penyebaran Kaus Demam Berdarah memiliki 96 *record*, untuk data *training* 76 *record* dan data *testing* memiliki 20 *record*.

Tabel 4.2 Data Training

N O	Daerah/Bulan	KP	CH	KU	Terinfeksi DBD
1	Padang Maninjau/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
2	Padang Maninjau/Desember	Sangat Padat	Tinggi	Sangat Tinggi	Terinfeksi
3	Panigoran/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
4	Panigoran/Okttober	Sangat Padat	Menengah	Tinggi	Terinfeksi
5	Panigoran/November	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
6	Panigoran/Desember	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
7	Sidomulyo/Juli	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
8	Sidomulyo/ Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
9	Sidomulyo/ September	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
⋮	⋮	⋮	⋮	⋮	⋮
94	Bandar Selamat/ September	Sangat Padat	Menengah	Tinggi	Terinfeksi
95	Purworejo/April	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
96	Purworejo/Mei	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi

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

Tabel 4.3 Data testing

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
1	Padang Maninjau/ Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
2	Padang Maninjau/ Desember	Sangat Padat	Tinggi	Sangat Tinggi	Terinfeksi
3	Panigoran/ Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
4	Panigoran/ Okttober	Sangat Padat	Menengah	Tinggi	Terinfeksi
5	Panigoran/ November	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi

6	Panigoran/ Desember	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
7	Sidomulyo/ Juli	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
8	Sidomulyo/ Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
9	Sidomulyo/ September	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
10	Karang Anyar/ Juni	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
11	Karang Anyar/ Juli	Sangat Padat	Menengah	Normal	Tidak Terinfeksi
12	Karang Anyar/Agustus	Sangat Padat	Tinggi	Sangat Tinggi	Tidak Terinfeksi
13	Karang Anyar/ September	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
14	Padang Halaban/ September	Sangat Padat	Menengah	Tinggi	Terinfeksi
15	Padang Halaban/Oktobe r	Sangat Padat	Menengah	Sangat Tinggi	Tidak Terinfeksi
16	Aek Korsik/Januari	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
17	Aek Korsik/Februari	Sangat Padat	Menengah	Tinggi	Tidak Terinfeksi
18	Purworejo/ November	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi
19	Bandar Selamat/ September	Sangat Padat	Menengah	Tinggi	Terinfeksi
20	Aek Korsik/ Oktober	Sangat Padat	Menengah	Sangat Tinggi	Terinfeksi

4.2 Pengujian Software RapidMiner

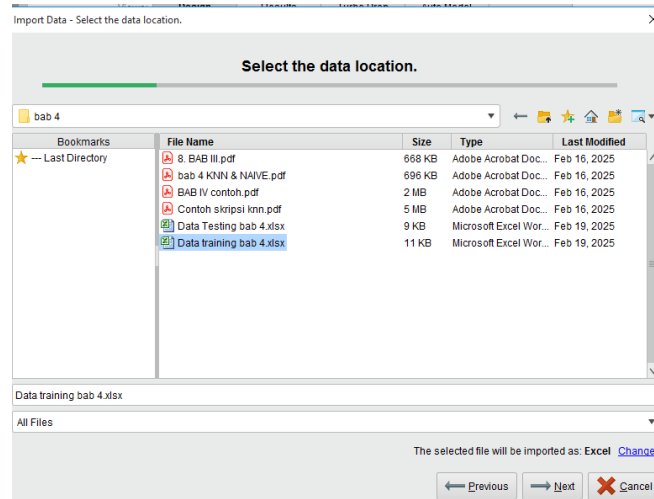
Pada tahap ini penulis mengambil data training 80% dari jumlah data keluruhan dan 20% data yang digunakan untuk data testing. Maka diketahui jika data keseluruhan tersebut berjumlah 96 data.

$$\text{Data training} = 80\% \times 96 = 76$$

$$\text{Data testing} = 20\% \times 96 = 20$$

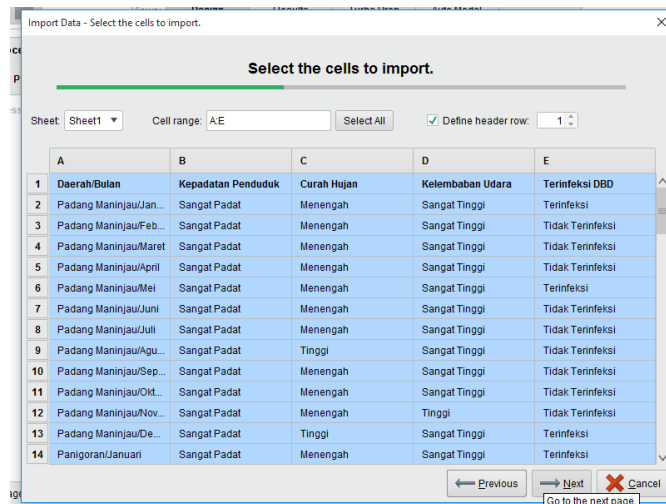
Atribut yang digunakan sebagai label adalah status *DBD*, sedangkan data *testing* yang digunakan diambil secara acak atau random dari 76 data yang telah

ada. Selanjutnya lakukan importing data terlebih dahulu yaitu *data training* dan *data testing* yang tersedia di pc atau computer seperti gambar 4.3



Gambar 4.3 Import data *training*

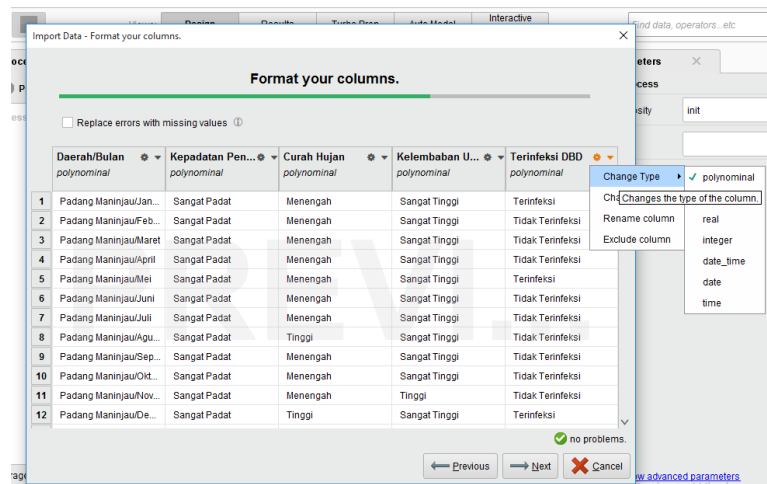
Setelah data *training* telah dipilih langkah selanjutnya yaitu klik *Next* seperti gambar 4.4 seperti gambar yang ada dibawah.



Gambar 4.4 Import Data *Training*

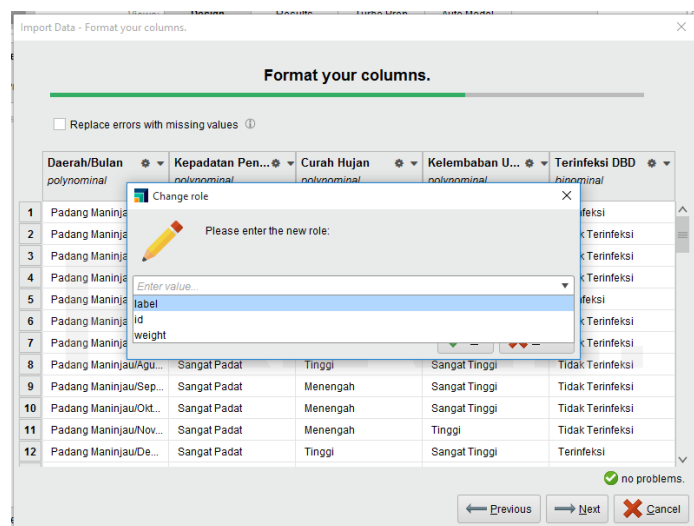
Untuk melanjutkan ke tahap selanjutnya klik *Next* maka akan muncul tampilan untuk memilih atau menyesuaikan atribut yang memiliki 2 kategori maka ia termasuk ke dalam binomial dan atribut yang memiliki lebih dari 2 kategori termasuk ke dalam polynomial. Pada gambar dibawah diketahui jika

yang termasuk kedalam binomial adalah *Status DBD*, dan atribut ini juga berupa label yang digunakan pada metode ini.



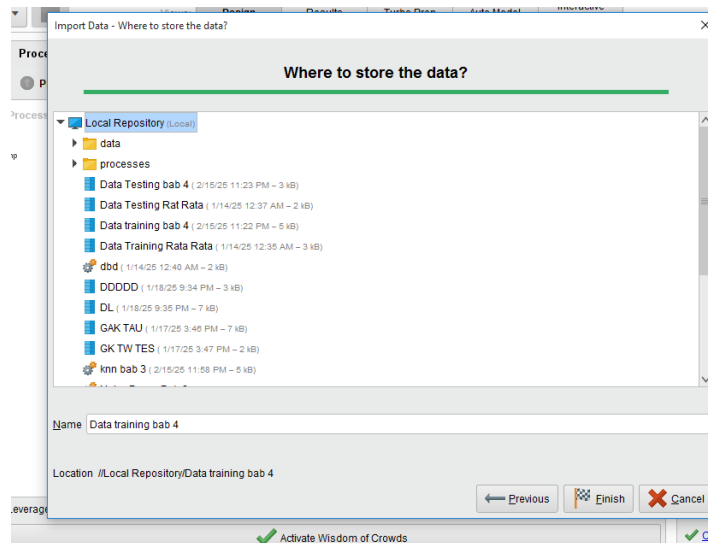
Gambar 4.5 Import data training

Selanjutnya kita pilih atribut yang digunakan sebagai label pada pengujian ini, selanjutnya bisa ubah pada change role dan pilih label lalu klik *Next*.



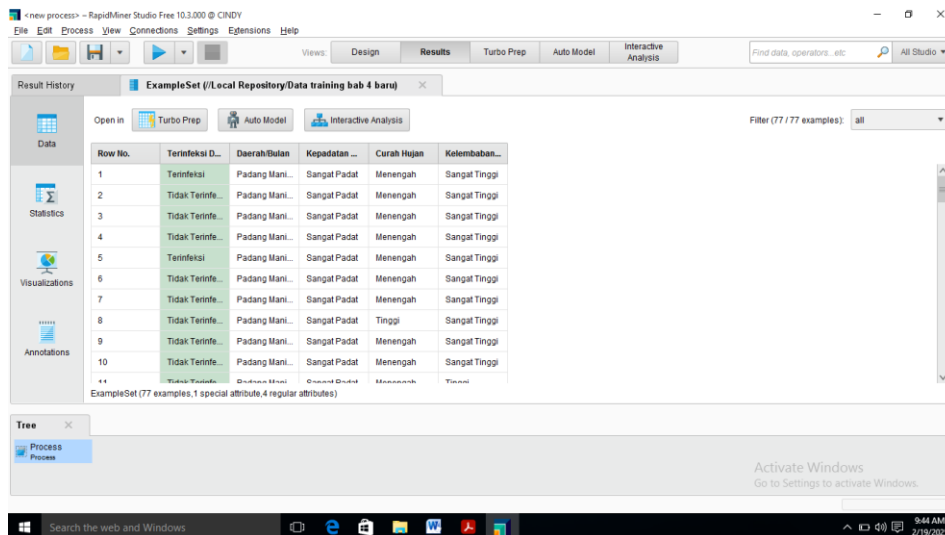
Gambar 4.6 Import data training

Setelah muncul tampilan seperti gambar 4.7 dibawah maka dapat kita klik *finish*



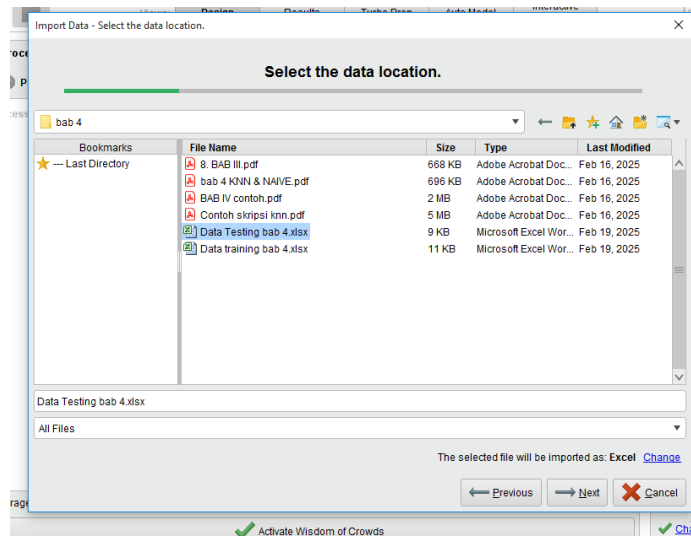
Gambar 4.7 Import data training

Jika sudah muncul tampilan seperti gambar 4.8 dibawah maka dapat diketahui jika data training yang kita *import* sudah berhasil.



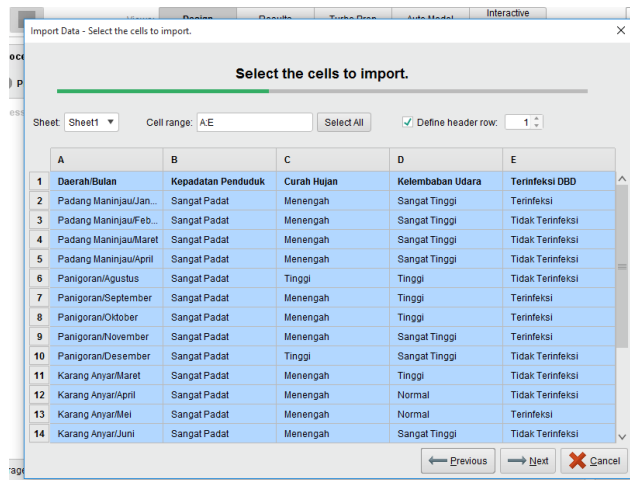
Gambar 4.8 Import Data Training

Langkah selanjutnya *import data testing* yang tersedia di pc atau komputer.



Gambar 4.9 Import data Testing

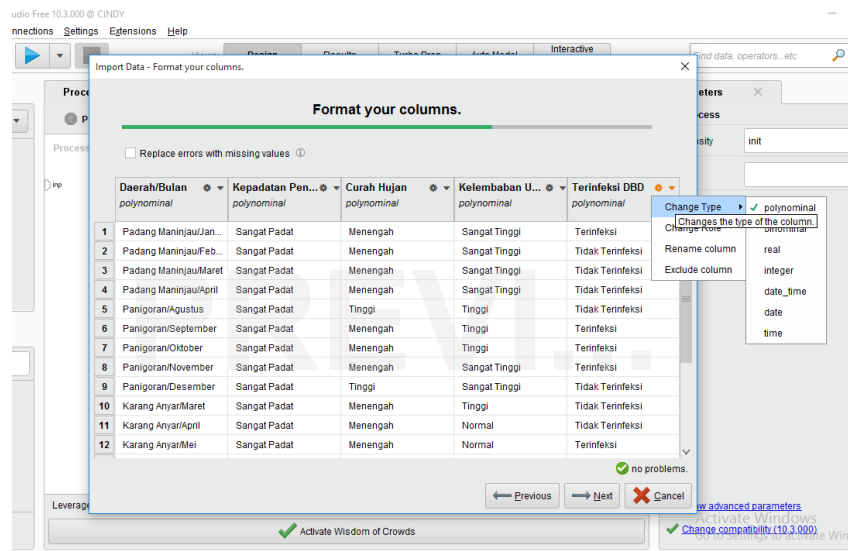
Setelah di pilih data atau file yang akan dijadikan data testing maka, dapat klik *Next* maka akan muncul form seperti gambar 4.10 dibawah ini.



Gambar 4.10 Import data testing

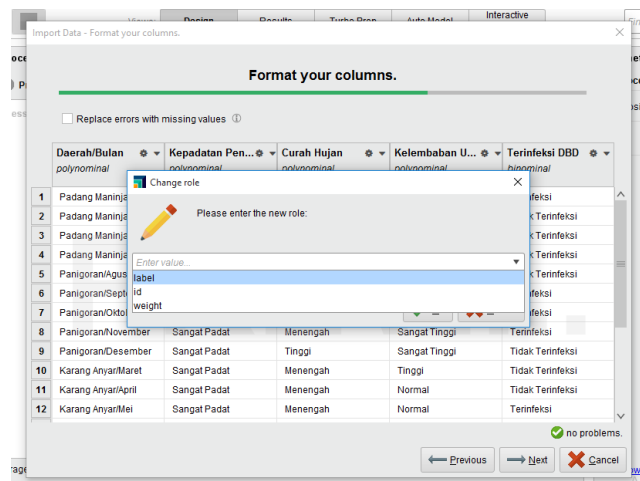
Untuk melanjutkan ketahap selanjutnya klik *Next* maka akan muncul form untuk menyesuaikan apakah kategori yang ada pada atribut tersebut termasuk kedalam binomial(yang terdiri dari 2 kategori) atau polynomial (terdiri lebih dari 2 kategori). Pada gambar 4.11 dibawah diketahui jika yang termasuk binomial

terdapat pada atribut status terinfeksi *DBD*. Maka bisa klik *change type* lalu pilih binomial.



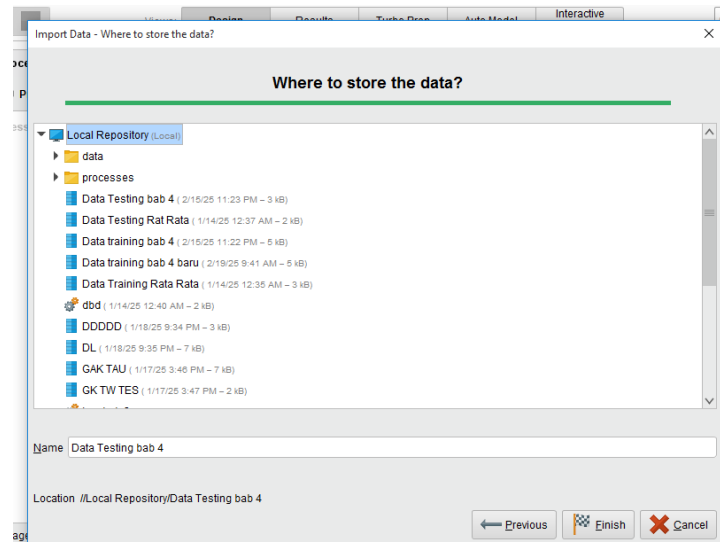
Gambar 4.11 Import data testing

Pada atribut *Status Terinfeksi DBD* merupakan atribut yang digunakan sebagai label. Maka klik pada *change role* lalu pilih label dan klik *ok* seperti gambar 4.12 yang ada dibawah ini.



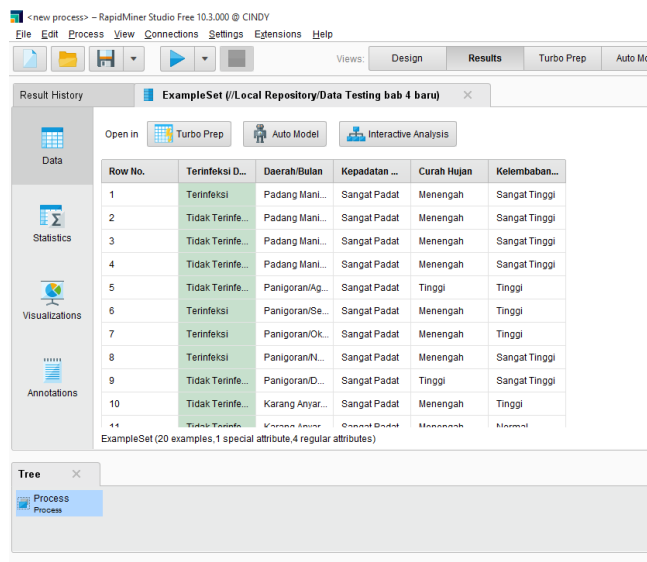
Gambar 4.12 Import data testing

Selanjutnya klik *Next* dan pilih tempat dimana file program yang dijalankan akan disimpan dimana, lalu klik *next*.



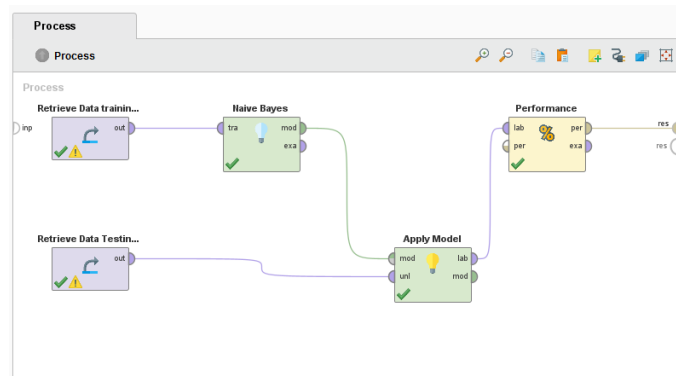
Gambar 4.13 Import data testing

Setelah selesai melakukan tahap pengimporan data testing maka klik *finish* dan jika muncul seperti gambar 4.14 dibawah ini



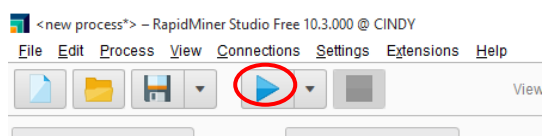
Gambar 4.14 Import data testing

Jika kedua data atau file sudah selesai di *import* maka langkah selanjutnya *drop* kedua file ke dalam lembar kerja dan drop *operator* *Naïve Bayes*, *Apply Model* dan *Performane* lalu hubungkan dengan data *training* dan data *testing* seperti gambar 4.15 dibawah ini



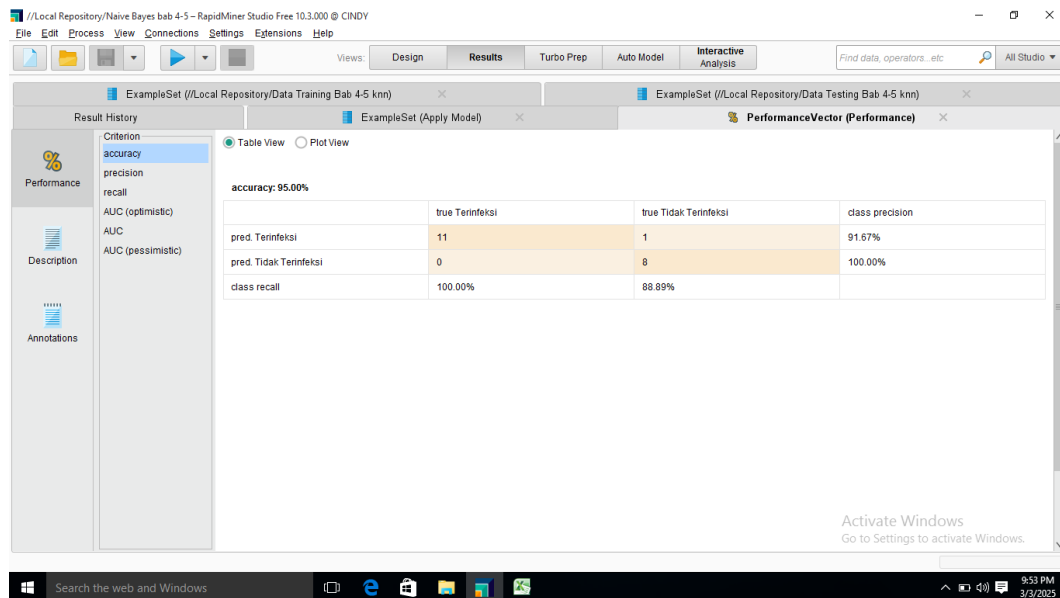
Gambar 4. 15 Menghubungkan operator

Setelah semua operator terhubung, selanjutnya klik icon *run* pada *toolbar* seperti gambar 4.16 dibawah ini untuk menampilkan hasil perhitungan *Naive Bayes*.



Gambar 4. 16 Icon menjalankan perintah

Selanjutnya *RapidMiner* akan menunjukkan hasil dari perhitungan yang telah dilakukan seperti gambar 4.17 dibawah ini



Gambar 4.17 Hasil dari perhitungan

Pada gambar 4.17 diatas tingkat *accuracy* dari *performace vector* adalah 95,00 %, *class precision Terinfeksi* 100,00% dan *class precision Tidak Terinfeksi* 88,89%, sedangkan untuk *class recall true Terinfeksi* 83,33% dan *class recall true Tidak Terinfeksi* 85,71%.

Secara umum precision, recall, dan accuracy dirumuskan sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+FP+FN+TN} \\
 &= \frac{11+8}{11+1+0+8} = \frac{19}{20} = 0.95 * 100\% = 95\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP+FP} \\
 &= \frac{11}{11+1} = \frac{11}{12} = 0.91 * 100\% = 91\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP+FN} \\
 &= \frac{TP}{TP+FN} = \frac{11}{11+0} = \frac{11}{11} = 1 * 100\% = 100\%
 \end{aligned}$$

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

$$= 2 \times \frac{0.91 \times 1}{0.91 + 1} = 2 \times \frac{0.91}{1.91} = 0.95$$

4.3 Pengujian Algoritma *Naïve Bayes* Pada *Microsoft Excel*

Pengujian pada penelitian ini menggunakan algoritma *Naïve Bayes* pada penelitian ini penulis membuat tabel data yang digunakan dalam penelitian seperti tabel 4.4 dan tabel 4.5. Untuk menghitung nilai *probabilitas* atau prediksi daerah yang berstatus terinfeksi atau tidak terinfeksi dengan menggunakan cara sebagai berikut:

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

1. Menghitung jumlah dari *Terinfeksi* dan *Tidak Terinfeksi* dari tabel data *Training*. Dari data *training I* didapatkan hasil sebagai berikut:

Terinfeksi = 19

Tidak Terinfeksi = 58

Setelah mengetahui jumlah status terinfeksi *DBD* yang terupa terinfeksi dan tidak terinfeksi selanjutnya adalah menghitung masing masing atribut dari data training seperti berikut:

Tabel 4.4 variabel status *DBD*

Variabel		P(Terinfeksi/Tidak Terinfeksi)
Status	Terinfeksi	21/77= 0.2727272723
DBD	Tidak Terinfeksi	58/77= 0.727272727
	TOTAL	100%

Menghitung setiap probabilitas variabel atau atribut pada setiap kelas

1. $P(\text{Daerah/Bulan} | \text{Status DBD (Terinfeksi \& Tidak Terinfeksi)})$

$$P(\text{terinfeksi}) = \frac{1}{21} = 0.04761904762$$

(Perhitungan diatas ialah jumlah data Daerah/Bulan “Padang Maninjau/Januari” dibagi dengan jumlah “Terinfeksi”)

$$P(\text{Tidak Terinfeksi}) = 0/21 = 0$$

(Perhitungan diatas ialah jumlah data Derah/Bulan “Padang Maninjau/Januari” Dibagi dengan jumlah “Tidak Terinfeksi”)

Tabel 4.5 Variabel Derah/Bulan

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
	Padang Maninjau/Januari	1	0	0.047619048	0
	Padang Maninjau/Februari	0	1	0	0.017857143
	Padang Maninjau/Maret	0	1	0	0.017857143
	Padang Maninjau/April	0	1	0	0.017857143
	Padang Maninjau/Mei	1	0	0.047619048	0
	Padang Maninjau/Juni	0	1	0	0.017857143
	Padang Maninjau/Juli	0	1	0	0.017857143
	Padang Maninjau/Agustus	0	1	0	0.017857143
	Padang Maninjau/September	0	1	0	0.017857143
	Padang Maninjau/Oktober	0	1	0	0.017857143
	Padang Maninjau/November	0	1	0	0.017857143
	Padang Maninjau/Desember	1	0	0.047619048	0
	Panigoran/Januari	1	0	0.047619048	0
	Panigoran/Februari	0	1	0	0.017857143
	Panigoran/Maret	0	1	0	0.017857143
	Panigoran/April	0	1	0	0.017857143
	Panigoran/Mei	0	1	0	0.017857143
	Panigoran/Juni	0	1	0	0.017857143
	Panigoran/Juli	0	1	0	0.017857143
	Panigoran/Agustus	0	1	0	0.017857143

	Panigoran/September	1	0	0.047619048	0
	Panigoran/Okttober	1	0	0.047619048	0
	Panigoran/November	1	0	0.047619048	0
	Panigoran/Desember	0	1	0	0.017857143
	Sidomulyo/Januari	0	1	0	0.017857143
	Sidomulyo/Februari	0	1	0	0.017857143
Daerah/	Sidomulyo/Mart	0	1	0	0.017857143
Bulan	Sidomulyo/April	0	1	0	0.017857143
	Sidomulyo/Mei	0	1	0	0.017857143
	Sidomulyo/Juni	0	1	0	0.017857143
	Sidomulyo/Juli	0	1	0	0.017857143
	Sidomulyo/Agustus	0	1	0	0.017857143
	Sidomulyo/September	1	0	0.047619048	0
	Sidomulyo/Okttober	1	0	0.047619048	0
	Sidomulyo/November	1	0	0.047619048	0
	Sidomulyo/Desember	0	1	0	0.017857143
	Karang Anyar/Januari	1	0	0.047619048	0
	Karang Anyar/Februari	0	1	0	0.017857143
	Karang Anyar/Maret	0	1	0	0.017857143
	Karang Anyar/April	0	1	0	0.017857143
	Karang Anyar/Mei	1	0	0.047619048	0
	Karang Anyar/Juni	0	1	0	0.017857143
	Karang Anyar/Juli	0	1	0	0.017857143
	Karang Anyar/Agustus	0	1	0	0.017857143
	Karang Anyar/September	1	0	0.047619048	0
	Karang Anyar/Okttober	0	1	0	0.017857143

	Karang Anyar/November	0	1	0	0.017857143
	Karang Anyar/Desember	0	1	0	0.017857143
	Padang Halaban/Januari	0	1	0	0.017857143
	Padang Halaban/Februari	0	1	0	0.017857143
	Padang Halaban/Maret	0	1	0	0.017857143
	Padang Halaban/April	0	1	0	0.017857143
	Padang Halaban/Mei	0	1	0	0.017857143
	Padang Halaban/Juni	0	1	0	0.017857143
	Padang Halaban/Juli	0	1	0	0.017857143
	Padang Halaban/Agustus	0	1	0	0.017857143
	Padang Halaban/September	1	0	0.047619048	0
	Padang Halaban/Okttober	0	1	0	0.017857143
	Padang Halaban/November	0	1	0	0.017857143
	Padang Halaban/Desember	0	1	0	0.017857143
	Aek Korsik/Januari	1	0	0.047619048	0
	Aek Korsik/Februari	0	1	0	0.017857143
	Aek Korsik/Maret	0	1	0	0.017857143
	Aek Korsik/April	0	1	0	0.017857143
	Aek Korsik/Mei	0	1	0	0.017857143
	Aek Korsik/Juni	0	1	0	0.017857143
	Aek Korsik/Juli	0	1	0	0.017857143
	Aek Korsik/Agustus	1	0	0.047619048	0
	Aek Korsik/September	1	0	0.047619048	0
	Aek Korsik/Okttober	1	0	0.047619048	0
	Aek Korsik/November	1	0	0.047619048	0
	Aek	0	1	0	0.017857143

	Korsik/Desember				3
	Purworejo/Januari	0	1	0	0.01785714 3
	Purworejo/November	1	0	0.047619048	0
	Bandar Selamat/September	1	0	0.047619048	0
	Purworejo/April	0	1	0	0.01785714 3
	Purworejo/Mei	0	1	0	0.01785714 3
		21	56	100%	100%

2. $P(\text{Kepadatan Penduduk}|\text{Status DBD (Terinfeksi \& Tidak Terinfeksi)})$

$$P(\text{Tidak Padat}|\text{Terinfeksi}) = 0/21 = 0$$

(Perhitungan diatas ialah jumlah data Kepadatan Penduduk “Tidak Padat” dibagi dengan jumlah “Terinfeksi”

$$P(\text{Tidak Padat}|\text{Tidak Terinfeksi}) = 0/56 = 0$$

Perhitungan diatas ialah jumlah data Kepadatan Penduduk “Tidak Padat” dibagi dengan jumlah “Tidak Terinfeksi”

Tabel 4.6 Variabel Kepadatan Penduduk

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
Kepadatan Penduduk	Tidak Padat	0	0	0	0
	Kurang Padat	0	0	0	0
	Cukup Padat	0	0	0	0
	Sangat Padat	21	56	1	1
	TOTAL	21	56	100%	100%

3. $P(\text{Curah Hujan}|\text{Status DBD (Terinfeksi \& Tidak Terinfeksi)})$

$$P(\text{Rendah}|\text{Terinfeksi}) = 0/21 = 0$$

Perhitungan diatas ialah jumlah data Curah Hujan “Rendah” dibagi dengan jumlah “Terinfeksi”

$$P(\text{Rendah}|\text{Tidak Terinfeksi}) = 0/21 = 0$$

Perhitungan diatas ialah jumlah data Curah Hujan “Rendah” dibagi dengan jumlah “Tidak Terinfeksi”

Tabel 4.7 Variabel Curah Hujan

Variabel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
	Rendah	0	0	0	0
Curah Hujan	Menengah	19	46	0.904761905	0.821428571
	Tinggi	2	10	0.095238095	0.178571429
	Sangat Tinggi	0	0	0	0
	TOTAL	21	56	100%	100%

4. $P(\text{Kelembaban Udara}|\text{Status DBD(Terinfeksi \& Tidak Terinfeksi)})$

$$P(\text{Sangat Rendah(Terinfeksi)}) = 0/21 = 0$$

Perhitungan diatas ialah jumlah data Kelembaban Udara “Sangat Rendah” dibagi dengan jumlah “Terinfeksi”

$$P(\text{Sangat Rendah(Tidak Terinfeksi)}) = 0/56 = 0$$

Perhitungan diatas ialah jumlah data Kelembaban Udara “Sangat Rendah” dibagi dengan jumlah “Tidak Terinfeksi”

Tabel 4.8 Variabel Kelembaban Udara

Varibel	Partisi	Terinfeksi	Tidak Terinfeksi	P(Terinfeksi)	P(Tidak Terinfeksi)
	Sangat Rendah	0	0	0	0
	Redah	0	0	0	0
Kelembaban	Normal	1	2	0.047619048	0.035714286
Udara	Tinggi	4	7	0.19047619	0.125
	Sangat Tinggi	16	47	0.761904762	0.839285714
	TOTAL	21	56	100%	100%

Menghitung perkalian probabilitas prior dan probabilitas setiap variabel pada setiap kelompok atau pada setiap posterior yang ada pada data *testing*. Berikut ini merupakan pengolahan data atau bentuk perkalian dengan menggunakan data pertama diantaranya sebagai berikut:

$$\begin{aligned}
 P_{(Terinfeksi)} &= P(\text{Daerah/Bulan}|\text{Padang Maninjau/Januari}) \times P(\text{Kepadatan Penduduk}|\text{Sangat Padat}) \times P(\text{Curah Hujan}|\text{Menengah}) \times \\
 &\quad P(\text{Kelembaban Udara}|\text{Sangat Tinggi}) \times P(\text{Status DBD}|\text{Terinfeksi}) \\
 &= 0.047619048 \times 1 \times 0.904761905 \times 0.761904762 \times 0.27272727273 \\
 &= 0.008952499
 \end{aligned}$$

$$\begin{aligned}
 P_{(Tidak Terinfeksi)} &= P(\text{Derah/Bulan}|\text{Padang Maninjau/Januari}) \times P(\text{Kepadatan penduduk}|\text{Sangat Padat}) \times (\text{Curah Hujan}|\text{Menengah}) \times \\
 &\quad P(\text{Kelembaban Udara}|\text{Sangat Tinggi}) \\
 &= 0 \times 1 \times 0.821428571 \times 0.39285714 \times 0.7272727272 \\
 &= 0
 \end{aligned}$$

Tabel 4.9 Hasil Klasifikasi Menggunakan *Naïve Bayes*

Daerah/Bulan	Terinfeksi DBD	Prediksi	CLASS PREDICTION
Padang Maninjau/Januari	Terinfeksi		TERINFEKSI
Padang Maninjau/Desember	Terinfeksi		TERINFEKSI
Panigoran/Januari	Terinfeksi		TERINFEKSI
Panigoran/Okttober	Terinfeksi		TERINFEKSI
Panigoran/November	Terinfeksi		TERINFEKSI
Panigoran/Desember	Tidak Terinfeksi		TIDAK TERINFEKSI
Sidomulyo/Juli	Tidak Terinfeksi		TIDAK TERINFEKSI
Sidomulyo/Agustus	Tidak Terinfeksi		TIDAK TERINFEKSI
Sidomulyo/September	Terinfeksi		TERINFEKSI
Karang Anyar/Juni	Tidak Terinfeksi		TIDAK TERINFEKSI
Karang Anyar/Juli	Tidak Terinfeksi		TIDAK TERINFEKSI
Karang Anyar/Agustus	Tidak Terinfeksi		TIDAK TERINFEKSI
Karang Anyar/September	Tidak Terinfeksi		TERINFEKSI
Padang Halaban/September	Terinfeksi		TERINFEKSI
Padang Halaban/Okttober	Tidak Terinfeksi		TIDAK TERINFEKSI
Aek Korsik/Januari	Terinfeksi		TERINFEKSI
Aek Korsik/Februari	Tidak Terinfeksi		TIDAK TERINFEKSI

Purworejo/November	Terinfeksi		TERINFEKSI
Bandar Selamat/September	Terinfeksi		TERINFEKSI
Aek Korsik/Oktober	Terinfeksi		TERINFEKSI

Pada tabel diatas merupakan hasil perhitungan yang dilakukan secara manual dengan menggunakan *Microsoft Excel* dengan mencantumkan hasil prediksi pada data uji (*testing*). Hasil klasifikasi metode *Naïve Bayes* menunjukkan bahwa dari 20 data uji ada 11 wilayah yang dinyatakan berisiko tinggi terkena penyakit *DBD* dan 9 wilayah yang dinyatakan tidak terinfeksi *DBD*. Maka dari itu dari hasil perhitungan prediksi yang telah dilakukan dapat dinyatakan jika Daerah Padang Maninjau memiliki kepadatan penduduk 4,702 km² dan Curah Hujan sebanyak 112 mm, serta Kelembaban Udara sebesar 70% maka dinyatakan berisiko tinggi terkena penyakit *DBD* atau terinfeksi. Sedangkan jika daerah Sidomulyo memiliki kepadatan penduduk sebanyak 1,907 km² dengan Curah Hujan 150 mm serta Kelembaban Udara sebesar 69% dinyatakan tidak berisiko terkena penyakit *DBD* atau terinfeksi.

4.4 Penerapan Metode *K-Nearest Neighbor*

Adapun langkah yang dapat dilakukan dalam penerepan metode *K-Nearest Neighbor* diantaranya adalah sebagai berikut.

4.4.1 Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data sekunder. Data sekunder sendiri merupakan sumber data yang didapatkan oleh peneliti baik melalui media perantara atau tidak secara langsung. Data sekunder yang diperoleh

dari Dinas Kesehatan Kec. Aek Kuo yaitu data jumlah kasus terinfeksi *DBD* perbulan selama 1 tahun terakhir (2023) yang terdiri dari 8 desa yang ada di Kec. Aek Kuo. Sedangkan data sekunder yang diperoleh dari Badan Pusat Statistika Labuhanbatu Utara yaitu data Kepadatan Penduduk di setiap desa, Curah Hujan dan Kelembaban Udara. Maka dapat diperoleh jumlah data yang digunakan sebanyak 96 data.

Tabel 4.4.1 View Data

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
1	Padang Maninjau/Januari	4,702	112	70	Terinfeksi
2	Padang Maninjau/Februari	4,702	280	72	Tidak Terinfeksi
3	Padang Maninjau/Maret	4,702	142	80	Tidak Terinfeksi
4	Padang Maninjau/April	4,702	101	80	Tidak Terinfeksi
5	Padang Maninjau/Mei	4,702	258	72.6	Terinfeksi
6	Padang Maninjau/Juni	4,702	140	77	Tidak Terinfeksi
7	Padang Maninjau/Juli	4,702	150	85	Tidak Terinfeksi
8	Padang Maninjau/Agustus	4,702	471	73	Tidak Terinfeksi
9	Padang Maninjau/September	4,702	153	93	Tidak Terinfeksi
10	Padang Maninjau/Okttober	4,702	185	92	Tidak Terinfeksi
11	Padang Maninjau/November	4,702	194	72	Tidak Terinfeksi
12	Padang Maninjau/Desember	4,702	333	89	Terinfeksi
13	Panigoran/Januari	1,570	112	80	Terinfeksi
14	Panigoran/Februari	1,570	280	80.2	Tidak Terinfeksi
15	Panigoran/Maret	1,570	142	78.4	Tidak Terinfeksi
16	Panigoran/April	1,570	101	78.4	Tidak Terinfeksi
17	Panigoran/Mei	1,570	258	73.8	Tidak Terinfeksi

18	Panigoran/Juni	1,570	140	71.9	Tidak Terinfeksi
19	Panigoran/Juli	1,570	150	68.5	Tidak Terinfeksi
20	Panigoran/Agustus	1,570	471	68.6	Tidak Terinfeksi
21	Panigoran/September	1,570	153	68.7	Tidak Terinfeksi
22	Panigoran/Oktober	1,570	185	69	Tidak Terinfeksi
23	Panigoran/November	1,570	194	71	Tidak Terinfeksi
24	Panigoran/Desember	1,570	333	79.1	Tidak Terinfeksi
25	Sidomulyo/Januari	1,907	112	89.3	Tidak Terinfeksi
26	Sidomulyo/Februari	1,907	280	85.8	Tidak Terinfeksi
27	Sidomulyo/Mart	1,907	142	87.8	Tidak Terinfeksi
28	Sidomulyo/April	1,907	101	86	Tidak Terinfeksi
29	Sidomulyo/Mei	1,907	258	81.3	Tidak Terinfeksi
30	Sidomulyo/Juni	1,907	140	79.3	Tidak Terinfeksi
31	Sidomulyo/Juli	1,907	150	76.5	Tidak Terinfeksi
32	Sidomulyo/Agustus	1,907	471	76.5	Tidak Terinfeksi
33	Sidomulyo/September	1,907	153	77.8	Tidak Terinfeksi
34	Sidomulyo/Oktober	1,907	185	75.3	Tidak Terinfeksi
35	Sidomulyo/November	1,907	194	77	Tidak Terinfeksi
36	Sidomulyo/Desember	1,907	333	92.3	Tidak Terinfeksi
37	Karang Anyar/Januari	1,036	112	65	Tidak Terinfeksi
38	Karang Anyar/Februari	1,036	280	77	Tidak Terinfeksi
39	Karang Anyar/Maret	1,036	142	77	Tidak Terinfeksi
40	Karang Anyar/April	1,036	101	65	Tidak Terinfeksi
41	Karang Anyar/Mei	1,036	258	65	Tidak Terinfeksi
42	Karang Anyar/Juni	1,036	140	72	Tidak Terinfeksi
43	Karang Anyar/Juli	1,036	150	68	Tidak Terinfeksi
44	Karang	1,036	471	98	Tidak Terinfeksi

	Anyar/Agustus				Terinfeksi
45	Karang Anyar/September	1,036	153	76	Terinfeksi
46	Karang Anyar/Oktober	1,036	185	70	Tidak Terinfeksi
47	Karang Anyar/November	1,036	194	67	Tidak Terinfeksi
48	Karang Anyar/Desember	1,036	333	98	Tidak Terinfeksi
49	Padang Halaban/Januari	1,500	112	82	Tidak Terinfeksi
50	Padang Halaban/Februari	1,500	280	83	Tidak Terinfeksi
51	Padang Halaban/Maret	1,500	142	78	Tidak Terinfeksi
52	Padang Halaban/April	1,500	101	78	Tidak Terinfeksi
53	Padang Halaban/Mei	1,500	258	75	Tidak Terinfeksi
54	Padang Halaban/Juni	1,500	140	74	Tidak Terinfeksi
55	Padang Halaban/Juli	1,500	150	71	Tidak Terinfeksi
56	Padang Halaban/Agustus	1,500	471	70	Tidak Terinfeksi
57	Padang Halaban/September	1,500	153	70	Terinfeksi
58	Padang Halaban/Oktober	1,500	185	73	Tidak Terinfeksi
59	Padang Halaban/November	1,500	194	75	Tidak Terinfeksi
60	Padang Halaban/Desember	1,500	333	74	Tidak Terinfeksi
61	Aek Korsik/Januari	15,976	112	70	Terinfeksi
62	Aek Korsik/Februari	15,976	280	62	Tidak Terinfeksi
63	Aek Korsik/Maret	15,976	142	87	Tidak Terinfeksi
64	Aek Korsik/April	15,976	101	80	Tidak Terinfeksi
65	Aek Korsik/Mei	15,976	258	65	Tidak Terinfeksi
66	Aek Korsik/Juni	15,976	140	59.2	Tidak Terinfeksi
67	Aek Korsik/Juli	15,976	150	78	Tidak Terinfeksi
68	Aek Korsik/Agustus	15,976	471	89	Terinfeksi
69	Aek Korsik/September	15,976	153	90	Terinfeksi

70	Aek Korsik/Oktobre	15,976	185	89	Terinfeksi
71	Aek Korsik/November	15,976	194	85	Terinfeksi
72	Aek Korsik/Desember	15,976	333	92	Tidak Terinfeksi
73	Purworejo/Januari	1,570	112	81.8	Tidak Terinfeksi
74	Purworejo/Februari	1,570	280	83	Tidak Terinfeksi
75	Purworejo/Maret	1,570	142	78.3	Tidak Terinfeksi
76	Purworejo/April	1,570	101	76.1	Tidak Terinfeksi
77	Purworejo/Mei	1,570	258	76	Tidak Terinfeksi
78	Purworejo/Juni	1,570	140	79	Tidak Terinfeksi
79	Purworejo/Juli	1,570	150	72.8	Tidak Terinfeksi
80	Purworejo/Agustus	1,570	471	72.1	Tidak Terinfeksi
81	Purworejo/September	1,570	153	72.7	Terinfeksi
82	Purworejo/Oktobre	1,570	185	73.5	Terinfeksi
83	Purworejo/November	1,570	194	76.3	Terinfeksi
84	Purworejo/Desember	1,570	333	79.6	Tidak Terinfeksi
85	Bandar Selamat/Januari	6,423	112	75	Tidak Terinfeksi
86	Bandar Selamat/Februari	6,423	280	81	Tidak Terinfeksi
87	Bandar Selamat/Maret	6,423	142	80	Tidak Terinfeksi
88	Bandar Selamat/April	6,423	101	75	Tidak Terinfeksi
89	Bandar Selamat/Mei	6,423	258	76	Tidak Terinfeksi
90	Bandar Selamat/Juni	6,423	140	77	Tidak Terinfeksi
91	Bandar Selamat/Juli	6,423	150	72	Tidak Terinfeksi
92	Bandar Selamat/Agustus	6,423	471	72	Tidak Terinfeksi
93	Bandar Selamat/September	6,423	153	67	Terinfeksi
94	Bandar Selamat/Oktobre	6,423	185	69	Terinfeksi
95	Bandar Selamat/November	6,423	194	76	Tidak Terinfeksi

96	Bandar Selamat/Desember	6,423	333	73	Tidak Terinfeksi
----	----------------------------	-------	-----	----	---------------------

Data asli yang didapatkan peneliti memiliki 4 atribut diantaranya yaitu daerah/bulan, kepadatan penduduk, curah hujan, kelembaban udara dan status terinfeksi.

4.4.2 Pembersihan Data

Setelah tahap pengumpulan data dan filter data maka tahap selanjutnya yaitu cleaning data agar tidak ada duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data seperti kesalahan cetak, sehingga data tersebut dapat diolah dan dilakukan proses data mining. Semua data yang dibutuhkan telah melalui tahap cleaning maka penulis mendapatkan 96 data yang akan diolah menjadi 2 bagian yaitu *data training* dan *data testing* yang akan dimasukkan atau diolah melalui bantuan *software* yang digunakan pada penelitian ini.

4.4.3 Implementasi Dengan K- Nearest Neighbor

Pada proses klasifikasi metode *K-NN* terdapat tiga alur yaitu menentukan nilai parameter *K*, menghitung jarak (includean) antara data training dan data testing, kemudian menentukan ranking dari hasil perhitungan jarak. Data yang digunakan pada metode *K-NN* yakni 80% untuk data training yang diambil dari jumlah keseluruhan data dan 20% untuk data testing yang diambil dari jumlah keseluruhan data.

Tabel 4.5.1 Data Trainingg K-NN

NO	Daerah/Bulan	KP	CH	KU	Terinfeksi DBD
1	Padang Maninjau/Januari	4,702	112	70	Terinfeksi
2	Padang Maninjau/Februari	4,702	280	72	Tidak Terinfeksi
3	Padang Maninjau/Maret	4,702	142	80	Tidak Terinfeksi
4	Padang Maninjau/April	4,702	101	80	Tidak Terinfeksi
5	Padang Maninjau/Mei	4,702	258	72.6	Terinfeksi
6	Padang Maninjau/Juni	4,702	140	77	Tidak Terinfeksi
7	Padang Maninjau/Juli	4,702	150	85	Tidak Terinfeksi
8	Padang Maninjau/Agustus	4,702	471	73	Tidak Terinfeksi
9	Padang Maninjau/September	4,702	153	93	Tidak Terinfeksi
10	Padang Maninjau/Okttober	4,702	185	92	Tidak Terinfeksi
11	Padang Maninjau/November	4,702	194	72	Tidak Terinfeksi
12	Padang Maninjau/Desember	4,702	333	89	Terinfeksi
13	Panigoran/Januari	1,570	112	80	Terinfeksi
14	Panigoran/Februari	1,570	280	80.2	Tidak Terinfeksi
15	Panigoran/Maret	1,570	142	78.4	Tidak Terinfeksi
16	Panigoran/April	1,570	101	78.4	Tidak Terinfeksi
17	Panigoran/Mei	1,570	258	73.8	Tidak Terinfeksi
18	Panigoran/Juni	1,570	140	71.9	Tidak Terinfeksi
19	Panigoran/Juli	1,570	150	68.5	Tidak Terinfeksi
20	Panigoran/Agustus	1,570	471	68.6	Tidak Terinfeksi
21	Panigoran/September	1,570	153	68.7	Terinfeksi
22	Panigoran/Okttober	1,570	185	69	Terinfeksi
23	Panigoran/November	1,570	194	71	Terinfeksi
24	Panigoran/Desember	1,570	333	79.1	Tidak Terinfeksi
25	Sidomulyo/Januari	1,907	112	89.3	Tidak Terinfeksi
26	Sidomulyo/Februari	1,907	280	85.8	Tidak Terinfeksi
27	Sidomulyo/Mart	1,907	142	87.8	Tidak Terinfeksi
28	Sidomulyo/April	1,907	101	86	Tidak Terinfeksi
29	Sidomulyo/Mei	1,907	258	81.3	Tidak Terinfeksi
30	Sidomulyo/Juni	1,907	140	79.3	Tidak Terinfeksi
31	Sidomulyo/Juli	1,907	150	76.5	Tidak Terinfeksi
32	Sidomulyo/Agustus	1,907	471	76.5	Tidak Terinfeksi
33	Sidomulyo/September	1,907	153	77.8	Terinfeksi
34	Sidomulyo/Okttober	1,907	185	75.3	Terinfeksi

35	Sidomulyo/November	1,907	194	77	Terinfeksi
36	Sidomulyo/Desember	1,907	333	92.3	Tidak Terinfeksi
37	Karang Anyar/Januari	1,036	112	65	Terinfeksi
38	Karang Anyar/Februari	1,036	280	77	Tidak Terinfeksi
39	Karang Anyar/Maret	1,036	142	77	Tidak Terinfeksi
40	Karang Anyar/April	1,036	101	65	Tidak Terinfeksi
41	Karang Anyar/Mei	1,036	258	65	Terinfeksi
42	Karang Anyar/Juni	1,036	140	72	Tidak Terinfeksi
43	Karang Anyar/Juli	1,036	150	68	Tidak Terinfeksi
44	Karang Anyar/Agustus	1,036	471	98	Tidak Terinfeksi
45	Karang Anyar/September	1,036	153	76	Terinfeksi
46	Karang Anyar/Oktober	1,036	185	70	Tidak Terinfeksi
47	Karang Anyar/November	1,036	194	67	Tidak Terinfeksi
48	Karang Anyar/Desember	1,036	333	98	Tidak Terinfeksi
49	Padang Halaban/Januari	1,500	112	82	Tidak Terinfeksi
50	Padang Halaban/Februari	1,500	280	83	Tidak Terinfeksi
51	Padang Halaban/Maret	1,500	142	78	Tidak Terinfeksi
52	Padang Halaban/April	1,500	101	78	Tidak Terinfeksi
53	Padang Halaban/Mei	1,500	258	75	Tidak Terinfeksi
54	Padang Halaban/Juni	1,500	140	74	Tidak Terinfeksi
55	Padang Halaban/Juli	1,500	150	71	Tidak Terinfeksi
56	Padang Halaban/Agustus	1,500	471	70	Tidak Terinfeksi
57	Padang Halaban/September	1,500	153	70	Terinfeksi
58	Padang Halaban/Oktober	1,500	185	73	Tidak Terinfeksi
59	Padang Halaban/November	1,500	194	75	Tidak Terinfeksi
60	Padang Halaban/Desember	1,500	333	74	Tidak Terinfeksi
61	Aek Korsik/Januari	15,976	112	70	Terinfeksi
62	Aek Korsik/Februari	15,976	280	62	Tidak Terinfeksi
63	Aek Korsik/Maret	15,976	142	87	Tidak Terinfeksi
64	Aek Korsik/April	15,976	101	80	Tidak Terinfeksi
65	Aek Korsik/Mei	15,976	258	65	Tidak Terinfeksi
66	Aek Korsik/Juni	15,976	140	59.2	Tidak Terinfeksi
67	Aek Korsik/Juli	15,976	150	78	Tidak Terinfeksi
68	Aek Korsik/Agustus	15,976	471	89	Terinfeksi
69	Aek Korsik/September	15,976	153	90	Terinfeksi
70	Aek Korsik/Oktober	15,976	185	89	Terinfeksi

71	Aek Korsik/November	15,976	194	85	Terinfeksi
72	Aek Korsik/Desember	15,976	333	92	Tidak Terinfeksi
73	Purworejo/Januari	1,570	112	81.8	Tidak Terinfeksi
74	Purworejo/November	1,570	194	76.3	Terinfeksi
75	Bandar Selamat/Desember	6,423	333	73	Tidak Terinfeksi
76	Purworejo/April	1,570	101	76.1	Tidak Terinfeksi
77	Purworejo/Mei	1,570	258	76	Tidak Terinfeksi

Ket:

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

Tabel 4.5.2 Data Testing K-NN

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD
1	Padang Maninjau/Januari	4,702	112	70	?
2	Padang Maninjau/Desember	4,702	333	89	?
3	Panigoran/Januari	1,570	112	80	?
4	Panigoran/Okttober	1,570	185	69	?
5	Panigoran/November	1,570	194	71	?
6	Panigoran/Desember	1,570	333	79.1	?
7	Sidomulyo/Juli	1,907	150	76.5	?
8	Sidomulyo/Agustus	1,907	471	76.5	?
9	Sidomulyo/September	1,907	153	77.8	?
10	Karang Anyar/Juni	1,036	140	72	?
11	Karang Anyar/Juli	1,036	150	68	?
12	Karang Anyar/Agustus	1,036	471	98	?
13	Karang Anyar/September	1,036	153	76	?
14	Padang Halaban/September	1,500	153	70	?
15	Padang Halaban/Okttober	1,500	185	73	?

16	Aek Korsik/Januari	15,976	112	70	?
17	Aek Korsik/Februari	15,976	280	62	?
18	Purworejo/November	1,570	194	76.3	?
19	Bandar Selamat/September	6,423	153	67	?
20	Aek Korsik/Oktober	15,976	185	89	?

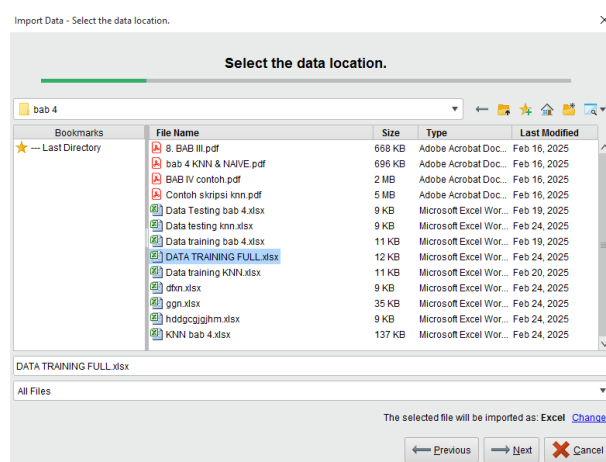
4.5 Pengujian Dengan Menggunakan *Software RapidMiner*

Pada tahap ini penulis mengambil data training 80% dari jumlah data keluruhan dan 20% data yang digunakan untuk data testing. Maka diketahui jika data keseluruhan tersebut berjumlah 96 data.

$$\text{Data training} = 80\% \times 96 = 77$$

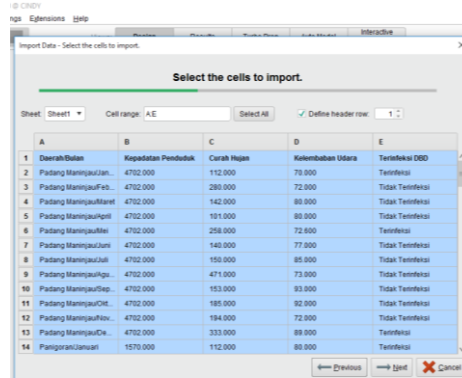
$$\text{Data testing} = 20\% \times 96 = 20$$

Atribut yang digunakan sebagai label adalah status *DBD*, sedangkan data *testing* yang digunakan diambil secara acak atau random dari 77 data yang telah ada. Selanjutnya lakukan importing data terlebih dahulu yaitu *data training* dan *data testing* yang tersedia di pc atau computer seperti gambar 4.17



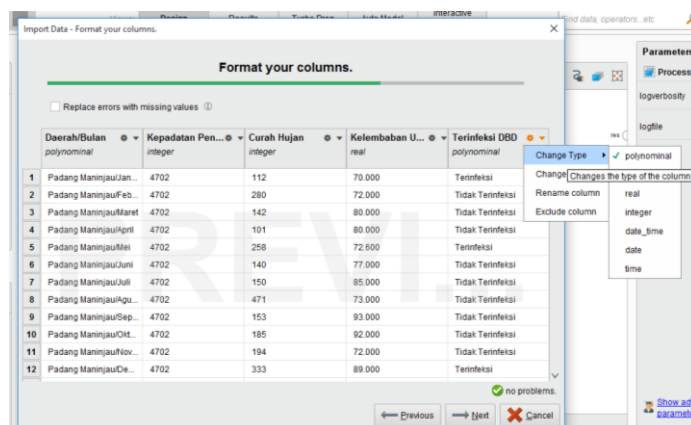
Gambar 4.18 Import data KNN

Setelah data dipilih langkah selanjutnya yaitu klik *Next* dan akan muncul form *data import* seperti gambar dibawah



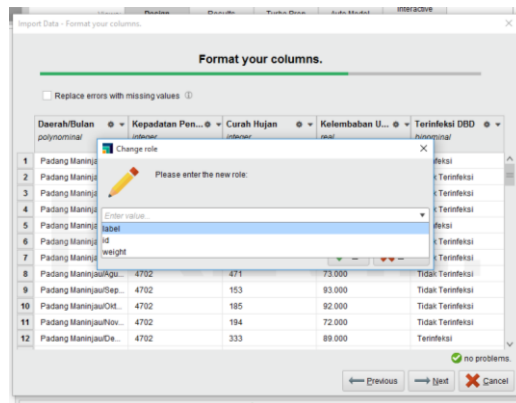
Gambar 4.19 Import data knn

Untuk melanjutkan ke tahap selanjutnya klik *Next* maka akan muncul tampilan untuk memilih atau menyesuaikan atribut yang memiliki 2 kategori maka ia termasuk ke dalam binomial dan atribut yang memiliki lebih dari 2 kategori termasuk ke dalam polynomial. Pada gambar dibawah diketahui jika yang termasuk kedalam binomial adalah *Status DBD*, dan atribut ini juga berupa label yang digunakan pada metode ini. Selanjutnya pilih *change type* lalu pilih binomial.



Gambar 4.19 Import data knn

Pada atribut terinfeksi *DBD* juga merupakan atribut yang digunakan sebagai label, maka dari itu bisa klik change role lalu pilih label.



Gambar 4.20 Import data knn

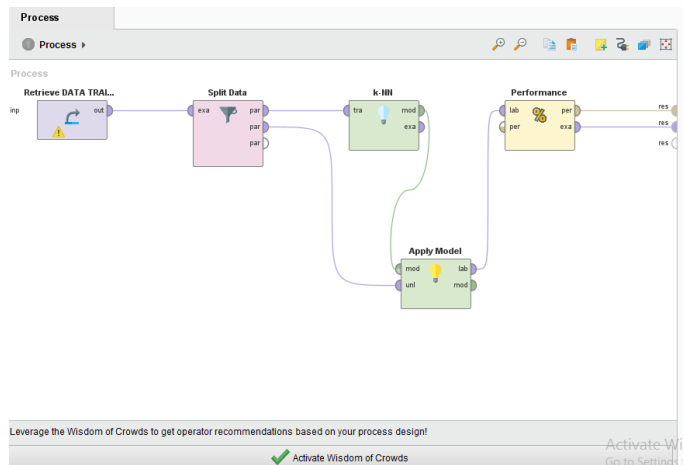
Selanjutnya klik *Next* dan pilih tempat penyimpanan yang sesuai. Jika sudah muncul tampilan seperti gambar dibawah maka itu berarti data yang di *import* sudah benar.

Row No.	Terinfeksi D...	Daerah/Bulan	Kepadatan ...	Curah Hujan	Kelembaban...
1	Terinfeksi	Padang Mani...	4702	112	70
2	Tidak Terife...	Padang Mani...	4702	280	72
3	Tidak Terife...	Padang Mani...	4702	142	80
4	Tidak Terife...	Padang Mani...	4702	101	80
5	Terinfeksi	Padang Mani...	4702	258	72.600
6	Tidak Terife...	Padang Mani...	4702	140	77
7	Tidak Terife...	Padang Mani...	4702	150	85
8	Tidak Terife...	Padang Mani...	4702	471	73
9	Tidak Terife...	Padang Mani...	4702	153	93
10	Tidak Terife...	Padang Mani...	4702	185	92
11	Tidak Terife...	Padang Mani...	4702	101	77

ExampleSet (96 examples, 1 special attribute, 4 regular attributes)

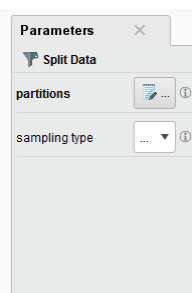
Gambar 4.21 Import data knn

Jika data atau file sudah selesai di *import* maka langkah selanjutnya *drop* kedua file ke dalam lembar kerja dan drop *operator* *K-NN*, *Apply Model*, *Split Data* dan *Performane* lalu hubungkan dengan data seperti gambar 4.23 dibawah ini.

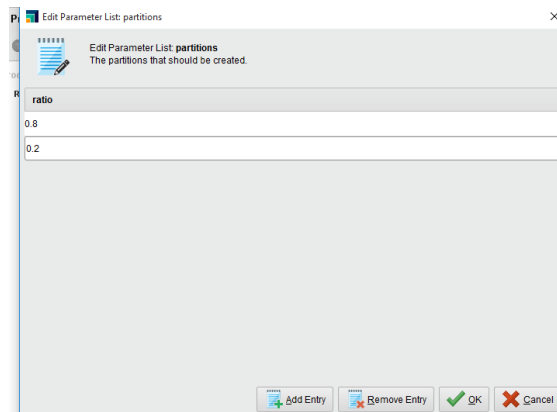


Gambar 4.22 Menghubungkan operator

Pada operator *split data* dapat dilakukan pembagian data sesuai dengan ketentuan yang telah ditetapkan yakni 80% untuk data training dan 20% untuk data *testingi* maka dapat dilakukan seperti klik klik dua kali pada operator split data, lalu para pamater pilih partitions dan masukkan atau input 0.8 (80%) untuk data training dan 0.2 (20%) untuk data testing seperti gambar 4.24 dan 4.25 dibawah

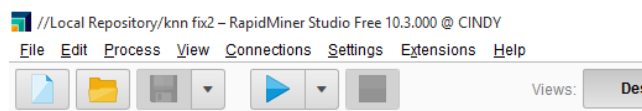


Gambar 4.24 import pembagian data



Gambar 4.25 import pembagian data training dan testing

Setelah semua operator terhubung, selanjutnya klik icon *run* pada *toolbar* seperti gambar 4.26 dibawah ini untuk menampilkan hasil perhitungan *k-NN*.



Gambar 4.26 icon *run*

Selanjutnya *RapidMiner* akan menunjukkan hasil dari perhitungan yang telah dilakukan seperti gambar 4.27 dibawah ini

//Local Repository/knn bab 4-5 re - RapidMiner Studio Free 10.3.000 @ CINDY

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Interactive Analysis

Result History: ExampleSet (Apply Model) PerformanceVector (Performance)

Open in: Turbo Prep Auto Model Interactive Analysis

Filter (20 / 20 examples): all

Row No.	Terinfeksi D...	prediction(T...	confidence(...	confidence(...	Daerah/Bulan	Kepadatan ...	Curah Hujan	Kelembaban...
1	Terinfeksi	Tidak Terinfe...	0.250	0.750	Padang Mani...	4702	112	70
2	Terinfeksi	Tidak Terinfe...	0.453	0.547	Padang Mani...	4702	333	89
3	Terinfeksi	Tidak Terinfe...	0.250	0.750	Panigoran/Ua...	1570	112	80
4	Terinfeksi	Terinfeksi	0.850	0.150	Panigoran/Ok...	1570	185	69
5	Terinfeksi	Terinfeksi	0.860	0.140	Panigoran/N...	1570	194	71
6	Tidak Terinfe...	Tidak Terinfe...	0	1	Panigoran/D...	1570	333	79.100
7	Tidak Terinfe...	Tidak Terinfe...	0.347	0.653	SidomulyoJu...	1907	150	76.500
8	Tidak Terinfe...	Tidak Terinfe...	0.166	0.834	SidomulyoAg...	1907	471	76.500
9	Terinfeksi	Tidak Terinfe...	0.374	0.626	SidomulyoS...	1907	153	77.800
10	Tidak Terinfe...	Tidak Terinfe...	0.319	0.681	Karang Anyar...	1036	140	72
11	Tidak Terinfe...	Tidak Terinfe...	0.218	0.782	Karang Anyar...	1036	150	68
12	Tidak Terinfe...	Tidak Terinfe...	0.185	0.815	Karang Anyar...	1036	471	98
13	Tidak Terinfe...	Tidak Terinfe...	0.250	0.750	Karang Anyar...	1036	153	76
14	Tidak Terinfe...	Tidak Terinfe...	0.250	0.750	Padang Hala...	1500	153	70
15	Terinfeksi	Tidak Terinfe...	0.183	0.817	Padang Hala...	1500	185	73

ExampleSet (20 examples, 4 special attributes, 4 regular attributes)

Activate Windows
Go to Settings to activate

Gambar 4.27 Hasil perhitungan jarak

	true Terinfeksi	true Tidak Terinfeksi	class precision
pred. Terinfeksi	4	0	100.00%
pred. Tidak Terinfeksi	7	9	56.25%
class recall	36.36%	100.00%	

Gambar 4.28 Hasil perhitungan klasifikasi

Pada gambar 4.28 diatas tingkat *accuracy* dari *performace vector* adalah 65,00 %, *class precision Terinfeksi* 100,00% dan *class precision Tidak Terinfeksi* 56,25%, sedangkan untuk *class recall true Terinfeksi* 36,36% dan *class recall true Tidak Terinfeksi* 100%.

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

$$= \frac{4+9}{4+0+7+9} = \frac{13}{20} = 0,65 * 100\% = 65\%$$

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

$$= \frac{4}{4+0} = 1 * 100\% = 100\%$$

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

$$= \frac{4}{4+7} = \frac{4}{11} = 0,36 * 100\% = 36\%$$

$$F1 Score = 2 * \frac{Precision * Recall}{Precision+Recall} = 2 * \frac{1*0,36}{1+0,36} = 2 * \frac{0,36}{1,36} = 0,52$$

4. 6 Pengujian Algoritma K- Nearest Neighbor Pada Microsoft Excel

Pengujian pada penelitian ini menggunakan algoritma *K-Nearest Neighbor*. Tahap pertama yang dapat dilakukan yakni perhitungan jarak

Eunclidean antara data training dengan data testing. Perhitungan jarak pada penelitian ini menggunakan data asli atau tanpa dilakukan transformasi terlebih dahulu. Adapun perhitungan jarak *encludiean* antara data training pertama dengan data testing pertama hingga data training ke -77 adalah sebagai berikut:

$$\begin{aligned}
& \sqrt{(X_1 - Y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \\
& \sqrt{(4702 - 4702)^2 + (112 - 112)^2 + (70 - 70)^2} \\
& \sqrt{0^2 + 0^2 + (0)^2} \\
& \sqrt{0} = 0 \\
\\
& \sqrt{(X_{77} - Y_1)^2 + (x_{78} - y_2)^2 + (x_{79} - y_3)^2} \\
& \sqrt{(4702 - 15.976)^2 + (112 - 185)^2 + (70 - 89)^2} \\
& \sqrt{(-11,274)^2 + (-73)^2 + (-19)^2} \\
& \sqrt{(12.103.076) + (5.329) + (361)} \\
& \sqrt{127.108.766} = 11.274,2523477
\end{aligned}$$

4.6.1 Menentukan Nilai Parameter K

Klasifikas Wilayah yang terinfeksi *DBD* dalam melihat apakah daerah tersebut terinfeksi atau tidak menggunakan metode *K-NN* dilakukan dengan menentukan nilai parameter K terlebih dahulu. Adapun nilai K yang digunakan yaitu K=5 dalam menghitung *Eucludien*. Berikut merupakan hasil perhitungan jarak antara data training pertama dengan data testing pertama hingga data training yang ke 77.

Tabel 4.5.3 Hasil perhitungan jarak dan $K=5$

Data Training ke-	Data testing ke-	Jarak (d)	Hasil $K=5$
1	1	0	1
2	1	168.0119	10
3	1	31.62278	4
4	1	14.86607	2
5	1	146.0231	9
6	1	28.86174	3
7	1	40.8534	5
8	1	359.0125	12
9	1	47.01064	6
10	1	76.24303	7
...	1		...
75	1	3132.155	32
76		3132.025	29

Pada tabel 4.17 diatas dilakukan perbandingan jarak antara semua data training dimana jumlah K yang telah ditentukan adalah 5, dimana setiap jarak *Euclidean* dari data training 1-77, jika jarak tersebut lebih kecil maka dilakukan pelabelan status *dbd* berdasarkan data testing terinfeksi atau tidak teinfeksi.

Tabel 4.5.4 Hasil nilai K

NO	Daerah/Bulan	Kepadatan Penduduk	Curah Hujan	Kelembaban Udara	Terinfeksi DBD	Encludiean
1	Padang Maninjau/Januari	4,702	112	70	Terinfeksi	0
4	Panigoran/April	1,570	101	78.4	Tidak Terinfeksi	14.86606875
6	Panigoran/Agustus	1,570	471	68.6	Tidak Terinfeksi	28.86173938
3	Padang Maninjau/Juli	4,702	150	85	Tidak Terinfeksi	31.6227766
7	Sidomulyo/Maret	1,907	142	87.8	Tidak Terinfeksi	40.85339643

Terlihat pada tabel diatas dengan menentukan $K=5$ atau 5 tetangga terdekat dari data atau nilai terkecil menunjukkan bahwa klasifikasi uji 1 sebanyak

5 *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 4.5.5 Hasil Klasifikasi

NO	Daerah/Bulan	Terinfeksi DBD	Prediksi	CLASS PREDICTION
1	Padang Maninjau/Januari	Terinfeksi		Tidak Terinfeksi
2	Padang Maninjau/Desember	Terinfeksi		Tidak Terinfeksi
3	Panigoran/Januari	Terinfeksi		Tidak Terinfeksi
4	Panigoran/Okttober	Terinfeksi		Terinfeksi
5	Panigoran/November	Terinfeksi		Terinfeksi
6	Panigoran/Desember	Tidak Terinfeksi		Tidak Terinfeksi
7	Sidomulyo/Juli	Tidak Terinfeksi		Tidak Terinfeksi
8	Sidomulyo/Agustus	Tidak Terinfeksi		Tidak Terinfeksi
9	Sidomulyo/September	Terinfeksi		Tidak Terinfeksi
10	Karang Anyar/Juni	Tidak Terinfeksi		Tidak Terinfeksi
11	Karang Anyar/Juli	Tidak Terinfeksi		Tidak Terinfeksi
12	Karang Anyar/Agustus	Tidak Terinfeksi		Tidak Terinfeksi
13	Karang Anyar/September	Tidak Terinfeksi		Tidak Terinfeksi
14	Padang Halaban/September	Tidak Terinfeksi		Tidak Terinfeksi
15	Padang Halaban/Okttober	Terinfeksi		Tidak Terinfeksi
16	Aek Korsik/Januari	Terinfeksi		Tidak Terinfeksi
17	Aek Korsik/Februari	Tidak Terinfeksi		Tidak Terinfeksi
18	Purworejo/November	Terinfeksi		Terinfeksi
19	Bandar Selamat/September	Terinfeksi		Tidak Terinfeksi
20	Aek Korsik/Okttober	Terinfeksi		Terinfeksi

Hasil klasifikasi dengan menggunakan metode *K-NN* menunjukkan bahwa dari 20 data uji ada 16 wilayah yang tidak terinfeksi *DBD*, 4 wilayah yang dinyatakan berisiko tinggi terkena penyakit *DBD* atau terinfeksi *DBD*. Maka dari itu dari hasil prediksi yang telah dilakukan dengan menggunakan metode *K-NN* dinyatakan jika Daerah Padang Maninjau dengan kepadatan Penduduk 4,702 km² dan Curah Hujan sebesar 112 mm serta Kelembaban Udara sebesar 70% dinyatakan tidak berisiko terkena atau terinfeksi *DBD*. Sedangkan pada daerah Aek Korsik dengan Kepadatan Penduduk sebanyak 15.976 km² dan Curah Hujan sebesar 101 serta dengan Kelembaban udara sebesar 76,1 dinyatakan berisiko tinggi terinfeksi *DBD*.

4.7 Pembahasan

Setelah dilakukan penelitian dengan menggunakan metode *Naïve Bayes* dalam memprediksi penyebaran *DBD* di kecamatan aek kuo dari 20 data uji dan 76 data training yang telah dilakukan pengujian terdapat 11 kategori berisiko tinggi terinfeksi *DBD* dengan melakukan perhitungan peluang ataupun probabilitas. Sedangkan pengujian yang telah dilakukan dengan menggunakan metode *K-NN* terdapat 4 kategori wilayah yang berisiko tinggi terinfeksi *DBD*. Penelitian yang telah dilakukan ini menghasilkan suatu informasi yang sesuai dengan tujuan data mining yaitu pola data *training* dan data *testing* untuk mendapatkan informasi baru, memprediksi daerah mana yang berisiko tinggi terinfeksi *DBD*. Proses klasifikasi menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbor*.

Setelah proses pengklasifikasian telah selesai maka selanjutnya proses pencarian tingkat akurasi dari kedua algoritma tersebut. Proses klasifikasi dan pencarian nilai akurasi menggunakan *software Rapidminer*. Dari proses tersebut pencarian nilai akurasi dari algoritma *Naïve Bayes* dan *K-Nearest Neighbor* maka diperoleh nilai akurasi *Navie bayes* sebesar 95%, Precision 91%, Recall 100% dan F1-Score sebesar 0,95 sedangkan *K-NN* sebesar 65,00 %, Precision 100%, Recall 36%, dan F1 Score 0,52 . Nilai akurasi membuktikan bahwa algoritma *K-NN* kurang baik digunakan untuk melakukan prediksi sedangkan nilai akurasi *Naïve Bayes* membuktikan bahwa *Naïve Bayes* sangat baik digunakan untuk memprediksi penyebaran daerah atau wilayah mana yang berisiko terinfeksi *DBD* di wilayah Kec. Aek Kuo.