

BAB III

METODELOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Subjek penelitian yang dilakukan dalam penelitian ini yaitu para pasien *COVID-19* diseluruh Indonesia dan India. Objek penelitian yang dilakukan yaitu klasifikasi status pasien *COVID-19* yang berupa data pasien.

3.2 Alat dan Bahan

Alat dan Bahan yang digunakan dalam penelitian ini antara lain:

3.2.1 Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

1. *Prosesor Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz (8 CPUs), ~2.3GHz*
2. *RAM 8GB*
3. *Storage 1TB HDD + Intel Optane 8GB*
4. *Display 15 Inch Full HD*
5. *Graphics Nvidia GeForce GTX 1050 Ti 4GB*

3.2.2 Perangkat Lunak

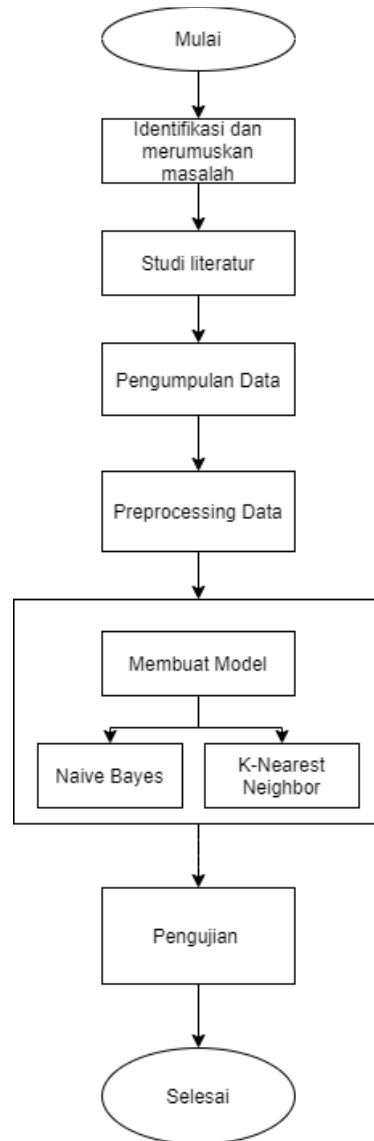
Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

1. *Windows 10 Home*
2. *Jupyter Notebook*
3. *Python 3.7+*
4. *Google Chrome*

3.2.3 Dataset

Dataset yang digunakan untuk penelitian ini adalah kumpulan dataset kaggle pasien corona Indonesia dan India.

3.3 Diagram Alur Penelitian



Gambar 3.1. Diagram alur penelitian

Untuk mendeskripsikan Gambar 3.1. diagram alur penelitian, maka peneliti membuat sub bab yang berisikan penjelasan mengenai tiap-tiap *chart* yang berada pada gambar tersebut. Berikut sub bab yang telah dibuat:

3.3.1 Identifikasi dan merumuskan masalah

Tahap ini mengidentifikasi masalah yang terjadi pada pandemi sekarang yaitu *COVID-19* seperti tingkat kematian yang mengacu dari data seperti umur, *gender*, jenis kelamin, serta kewarganegaraan yang banyak terjadi penularan *COVID-19*.

3.3.2 Studi literatur

Tahap ini peneliti melakukan studi literatur atau riset pada penelitian yang berhubungan dengan permasalahan serta metode yang dipakai, peneliti melakukan pengumpulan data-data sesuai atau berkaitan dengan permasalahan yaitu melakukan perbandingan klasifikasi status pasien *COVID-19* dan algoritma yang dipakai yaitu *Naïve bayes* dan *K-Nearest Neighbor*. Studi literatur yang disusun digunakan untuk memperkuat permasalahan yang dibahas pada penelitian ini.

3.3.3 Pengumpulan Data

Data yang peneliti gunakan dalam penelitian ini adalah dataset publik yang disediakan oleh *website* Kaggle, *download* dari web <https://www.kaggle.com/ardisragen/indonesiacoronaviruscases?select=patient.csv> untuk dataset Indonesia dan <https://www.kaggle.com/sudalairajkumar/covid19in-india/version/204?select=IndividualDetails.csv> untuk dataset India. Dataset tersebut antara lain sebagai berikut pada Tabel 3.1 dan Tabel 3.2.

Tabel 3.1. Contoh dataset pasien Indonesia

patient_id	gender	age	nationality	province	Currentstate	contacted_with	confirmed_date	released_date	deceased_date	hospital
1	female	31	indonesia	DKI Jakarta	released		2-Mar-20	13-Mar-20		RSPI Suliанти Saroso
2	female	64	indonesia	DKI Jakarta	released	1	2-Mar-20	16-Mar-20		RSPI Suliанти Saroso
3	female	33	indonesia	DKI Jakarta	released	1	6-Mar-20	13-Mar-20		RSPI Suliанти Saroso
4	female	34	indonesia	DKI Jakarta	isolated	1	6-Mar-20			RSPI Suliанти Saroso
5	male	55	indonesia	DKI Jakarta	isolated	1	8-Mar-20			RS Persahabatan
6	male	36	indonesia	DKI Jakarta	released		8-Mar-20	11-Mar-20		RS Persahabatan
7	female	54	indonesia	DKI Jakarta	released	8	9-Mar-20			
8	male	55	indonesia	DKI Jakarta	released		9-Mar-20			
9	female	55	indonesia	DKI Jakarta	released		9-Mar-20			
10	male	29	foreigner	DKI Jakarta	isolated	1	9-Mar-20			RSPI Suliанти Saroso

Tabel 3.2 Contoh dataset pasien India

id	gender	government_id	diagnosed_date	province	Currentstate	contacted_with	confirmed_date	released_date	deceased_date	hospital		
1	KL-AL-P1	KL-TS-P1	2/2/2020			Alappuzha	Alappuzha	Kerala	India	Recovered	14/02/2020	Travelled from Wuhan
2	KL-KS-P1	KL-AL-P1	3/2/2020			Kasaragod	Kasaragod	Kerala	India	Recovered	14/02/2020	Travelled from Wuhan
3	DL-P1	KL-KS-P1	2/3/2020	45	M	East Delhi (Mayur Vihar)	East Delhi	Delhi	India	Recovered	15/03/2020	Travelled from Austria, Italy
4	TS-P1	DL-P1	2/3/2020	24	M	Hyderabad	Hyderabad	Telangana	India	Recovered	2/3/2020	Travelled from Dubai to Bangalore on 20th Feb, stayed there for 2 days and took a bus to Hyderabad
5		TS-P1	3/3/2020	69	M	Jaipur	Jaipur	Rajasthan	Italy	Recovered	3/3/2020	Travelled from Italy
6			4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered	29/03/2020	Travelled from Italy

id	gender	government_id	diagnosed_date	province	Currentstate	contacted_with	confirmed_date	released_date	deceased_date	hospital		
7			4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered	29/03/2020	Travelled from Italy
8			4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered	29/03/2020	Travelled from Italy
9			4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered	29/03/2020	Travelled from Italy
10			4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered	29/03/2020	Travelled from Italy

3.3.4 Data Preprocessing

Pada tahap *preprocessing data*, data yang telah didapat dilakukan perubahan sedemikian rupa sehingga data yang telah diolah dapat dilakukan pengujian. Berikut adalah langkah-langkah data *preprocessing*:

1. Data Cleansing

Data cleansing bertujuan untuk membersihkan nilai kosong, palsu dan tidak konsisten (*missing value* dan *noisy*). Berikut merupakan hasil perbandingan data setelah melakukan *data cleaning*:

37	70 M	India	Recovered
38	68 F	India	Recovered
39		India	Hospitalized
40		India	Hospitalized
41		India	Recovered
42	27 M	India	Recovered
43	69 F	India	Recovered

Gambar 3.2. Contoh data yang belum dilakukan *cleansing*

Pada Gambar 3.2 terdapat nilai kosong, maka nilai yang kosong tersebut dilakukan pembersihan dengan cara menghapus baris pada data yang kosong secara manual.

37	70 M	India	Recovered
38	68 F	India	Recovered
42	27 M	India	Recovered
43	69 F	India	Recovered

Gambar 3.3. Contoh data yang sudah dilakukan *cleansing*

Pada Gambar 3.3 merupakan hasil data yang telah dilakukan *data cleaning*.

2. Data Integration

Data Integration tahap ini merupakan data yang awalnya terpisah disatukan agar data yang dihasilkan lebih baik. Berikut merupakan hasil perbandingan setelah melakukan *data integration*:

id	gender	age	nationality	province	current_status	contacted	confirmed	released	deceased	hospital	id	government	diagnosed	age	gender	detected	detected	detected	nationality	current_status	
1	female	31	indonesia	DKI Jakarta	released	2-Mar-20	#####			RSP	2	0	KL-TS-P1	30/01/202	20	F	Thirissur	Thirissur	Kerala	India	Recovered 1
2	female	64	indonesia	DKI Jakarta	released	1 2-Mar-20	#####			RSP	3	1	KL-AL-P1	2/2/2020			Alappuzha	Alappuzha	Kerala	India	Recovered 1
3	female	33	indonesia	DKI Jakarta	released	1 6-Mar-20	#####			RSP	4	2	KL-KS-P1	3/2/2020			Kasaragod	Kasaragod	Kerala	India	Recovered 1
4	female	34	indonesia	DKI Jakarta	isolated	1 6-Mar-20				RSP	5	3	DL-P1	2/3/2020	45	M	East Delhi	East Delhi	Delhi	India	Recovered 1
5	male	55	indonesia	DKI Jakarta	isolated	1 8-Mar-20				RS P	6	4	TS-P1	2/3/2020	24	M	Hyderabad	Hyderabad	Telangana	India	Recovered
6	male	36	indonesia	DKI Jakarta	released	8-Mar-20	#####			RS P	7	5		3/3/2020	69	M	Jaipur	Jaipur	Rajasthan	Italy	Recovered
7	female	54	indonesia	DKI Jakarta	released	8 9-Mar-20					8	6		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
8	male	55	indonesia	DKI Jakarta	released	9-Mar-20					9	7		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
9	female	55	indonesia	DKI Jakarta	released	9-Mar-20					10	8		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
10	male	29	foreigner	DKI Jakarta	isolated	1 9-Mar-20				RSP	11	9		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
11	female	54	foreigner	DKI Jakarta	isolated	1 9-Mar-20				RSP	12	10		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
12	male	31	indonesia	DKI Jakarta	isolated	1 9-Mar-20					13	11		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2
13	female	16	indonesia	DKI Jakarta	released	1 9-Mar-20	#####				14	12		4/3/2020	55		Gurugram	Gurugram	Haryana	Italy	Recovered 2

Gambar 3.4. Contoh data yang belum dilakukan *integration*

Pada Gambar 3.4 terdapat dua data yang berbeda, pada sebelah kiri merupakan data Indonesia dan sebelah kanan merupakan data India. Kedua data itu memiliki kesamaan atribut yang diteliti yaitu *age*, *nationality*, dan *current status*.

id	age	gender	nationality	current_status
1	31	female	indonesia	released
2	64	female	indonesia	released
3	33	female	indonesia	released
4	34	female	indonesia	isolated
5	55	male	indonesia	isolated
6	36	male	indonesia	released
7	54	female	indonesia	released
8	55	male	indonesia	released
0	20	F	India	released
1	45	M	India	released
2	24	M	India	released
3	69	M	India	released

Gambar 3.5. Contoh data yang sudah dilakukan *integration*

Pada Gambar 3.5 merupakan hasil data yang telah di *integration*.

3. Data Selection

Data Selection adalah proses menganalisis data-data yang digunakan dan dibutuhkan dalam melakukan data mining. Dari semua atribut yang digunakan

adalah *age*, *gender*, *nationality* dan *current status*. Atribut yang disebutkan sebelumnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penelitian. Berikut merupakan hasil perbandingan setelah melakukan *data selection*:

patient_id	gender	age	nationality	province	current_st	contacted	confirmed	released_c	deceased	hospital
1	female	31	indonesia	DKI Jakarta	released		2-Mar-20	#####		RSPI Sulianti Saroso
2	female	64	indonesia	DKI Jakarta	released	1	2-Mar-20	#####		RSPI Sulianti Saroso
3	female	33	indonesia	DKI Jakarta	released	1	6-Mar-20	#####		RSPI Sulianti Saroso
4	female	34	indonesia	DKI Jakarta	isolated	1	6-Mar-20			RSPI Sulianti Saroso
5	male	55	indonesia	DKI Jakarta	isolated	1	8-Mar-20			RS Persahabatan
6	male	36	indonesia	DKI Jakarta	released		8-Mar-20	#####		RS Persahabatan
7	female	54	indonesia	DKI Jakarta	released	8	9-Mar-20			
8	male	55	indonesia	DKI Jakarta	released		9-Mar-20			
9	female	55	indonesia	DKI Jakarta	released		9-Mar-20			
10	male	29	foreigner	DKI Jakarta	isolated	1	9-Mar-20			RSPI Sulianti Saroso
11	female	54	foreigner	DKI Jakarta	isolated	1	9-Mar-20			RSPI Sulianti Saroso

Gambar 3.6. Contoh data yang belum dilakukan *selection*

Pada Gambar 3.6 terdapat data dengan atribut yang tidak dibutuhkan seperti *province*, *contacted*, *confirmed*, *released date*, *deceased date*, dan *hospital*. Data atribut yang disebutkan dihapus karena tidak dibutuhkan dalam penelitian ini.

patient_id	gender	age	nationality	current_state
1	female	31	indonesia	released
2	female	64	indonesia	released
3	female	33	indonesia	released
4	female	34	indonesia	isolated
5	male	55	indonesia	isolated
6	male	36	indonesia	released
7	female	54	indonesia	released
8	male	55	indonesia	released
9	female	55	indonesia	released
10	male	29	foreigner	isolated
11	female	54	foreigner	isolated
12	male	31	indonesia	isolated
13	female	16	indonesia	released

Gambar 3.7. Contoh data yang sudah dilakukan *selection*

Pada gambar 3.7 merupakan hasil data yang telah *diselection*.

4. Data Transformation

Tahap selanjutnya dari proses *data cleaning* adalah *data transformation* dimana data yang diproses dilakukan transformasi dan inisialisasi agar penelitian dari data yang diproses dapat berjalan dengan baik. Berikut tabel inisialisasi:

Tabel 3.3. Inisialisasi Jenis Kelamin/*Gender*

Laki Laki	1
Perempuan	2

Tabel 3.4. Inisialisasi Kebangsaan/*Nationality*

Kebangsaan Indonesia/India	1
Kebangsaan Asing	2

Tabel 3.5. Inisialisasi Status Pasien

Released/Sembuh	0
Isolated/Isolasi	1
Deceased/Meninggal	2

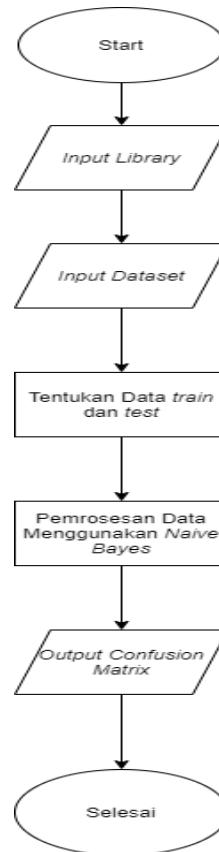
Tabel 3.6. Hasil Preprocessing Inisialisasi

age	gender	nationality	current_status
86	1	2	1
68	2	1	1
43	2	1	1
78	1	1	1
67	2	1	1
66	1	1	2
20	2	1	0
45	1	1	0
24	1	1	0
69	1	2	0
70	2	2	0
45	2	1	0
16	1	1	0

age	gender	nationality	current_status
27	1	1	1
59	1	1	0
65	1	1	1
75	1	1	1
45	1	1	0
24	1	1	0

3.3.5 Membuat Model

Tahap ini peneliti membuat model klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor* dan dilakukan pengujian dari dataset yang telah diambil dengan melakukan perbandingan dari kedua algoritma yaitu *Naïve Bayes* dan *K-Nearest Neighbor* yang diuji dengan menggunakan software yaitu *Python* berikut adalah cara kerja dari setiap algoritma.



Gambar 3.8. Cara kerja algoritma *Naïve Bayes*

Penjelasan cara kerja algoritma *Naïve Bayes* dari Gambar 3.8 yaitu yang pertama kali dilakukan adalah memulai, langkah selanjutnya adalah menginputkan *library* yang diperlukan seperti *numpy*, *pandas*, *matplotlib*, dan *sckit-learn*, selanjutnya menginputkan *dataset*, setelah itu menentukan data *train* dan *test*, setelah itu memproses dataset yang telah diinput dengan menggunakan algoritma *Naïve Bayes*, kemudian *output* yang didapatkan merupakan hasil evaluasi dari algoritma yaitu *confusion matrix*.



Gambar 3.9. Cara kerja algoritma *K-Nearest Neighbor*

Penjelasan cara kerja algoritma *K-Nearest Neighbor* dari Gambar 3.9 yaitu yang pertama kali dilakukan adalah memulai, kemudian langkah selanjutnya adalah menginputkan *library* yang diperlukan seperti *numpy*, *pandas*, *matplotlib*, dan *sckit-learn*, selanjutnya menginputkan *dataset*, setelah itu menentukan data *train* dan *test*, selanjutnya menentukan nilai tetangga terdekat atau nilai *K*, setelah itu memproses *dataset* yang telah *input* dan nilai *K* yang diberikan dengan menggunakan algoritma *K-Nearest Neighbor*, kemudian *output* yang didapatkan merupakan hasil evaluasi dari algoritma yaitu *confusion matrix*.

3.3.6 Pengujian

Setelah dilakukan implementasi, maka dilakukan pengujian dari data yang telah diambil menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor*. Tahap pengujian dilakukan menggunakan *Jupyter Notebook*. Pada pengujian ini dilakukan evaluasi untuk mencari akurasi, *precision* dan *recall* seperti ditunjukkan pada persamaan/rumus (3), (4), dan (5).

3.3.6.1 Perhitungan Naïve Bayes

Dataset yang digunakan sebagai data *training* adalah sebanyak 560 data sedangkan data *testing* yang ditentukan sebanyak 240 data. Berikut merupakan perhitungan *Naïve Bayes*:

1. Menghitung Probabilitas Kelas

Tahap pertama perhitungan dengan menggunakan *Naïve Bayes* adalah dengan mencari probabilitas dari masing masing kelas yaitu *released*, *isolated*, dan *deceased*. Hasil dari perhitungan tersebut dapat dilihat dari table berikut:

Table 3.7. Probabilitas Kelas

Kelas		
Released	Isolated	Deceased
77/800	707/800	16/800

2. Menghitung Probabilitas Masing-Masing Atribut

Untuk mencari probabilitas suatu atribut adalah dengan membandingkan atribut dari data *testing* dengan atribut data *training*.

1. Atribut Jenis Kelamin

Tabel 3.8. Atribut Jenis Kelamin

Jenis kelamin	Released	Isolated	Deceased
1(Laki-Laki)	50/800	615/800	13/800
2(Perempuan)	27/800	92/800	3/800

2. Atribut Umur

Tabel 3.9. Atribut Umur

Umur	Released	Isolated	Deceased
2	0/800	1/800	0/800
3	1/800	4/800	0/800
7	0/800	1/800	0/800
9	0/800	1/800	0/800
12	0/800	1/800	0/800
13	1/800	0/800	0/800
16	1/800	0/800	0/800
17	0/800	1/800	0/800
18	1/800	2/800	0/800
19	0/800	2/800	0/800
20	2/800	6/800	0/800
21	1/800	11/800	0/800
22	2/800	11/800	0/800
23	4/800	6/800	0/800
24	5/800	11/800	0/800
25	1/800	10/800	0/800
26	2/800	10/800	0/800
27	3/800	9/800	0/800
28	0/800	2/800	0/800
29	0/800	3/800	0/800
30	0/800	6/800	0/800
31	0/800	4/800	0/800
32	4/800	8/800	0/800

Umur	<i>Released</i>	<i>Isolated</i>	<i>Deceased</i>
33	2/800	8/800	0/800
34	1/800	5/800	0/800
35	4/800	6/800	0/800
36	1/800	5/800	0/800
37	1/800	5/800	0/800
38	1/800	5/800	1/800
39	0/800	4/800	0/800
40	2/800	3/800	0/800
41	2/800	7/800	0/800
42	0/800	1/800	0/800
43	0/800	4/800	0/800
44	0/800	2/800	0/800
45	4/800	456/800	0/800
46	2/800	2/800	0/800
47	2/800	4/800	0/800
48	2/800	1/800	0/800
49	0/800	7/800	0/800
50	0/800	2/800	1/800
51	1/800	2/800	0/800
52	0/800	5/800	0/800
53	2/800	3/800	0/800
54	1/800	2/800	1/800
55	0/800	4/800	0/800
56	2/800	2/800	0/800
57	0/800	5/800	1/800
58	0/800	4/800	0/800
59	2/800	4/800	0/800
60	0/800	4/800	0/800
61	1/800	3/800	0/800
62	0/800	1/800	0/800
63	3/800	4/800	1/800
64	1/800	4/800	0/800
65	3/800	4/800	2/800
66	0/800	1/800	1/800
67	1/800	2/800	1/800
68	0/800	4/800	0/800
69	2/800	3/800	3/800
70	2/800	1/800	2/800
73	1/800	0/800	0/800
74	0/800	2/800	0/800

Umur	Released	Isolated	Deceased
75	0/800	3/800	0/800
76	0/800	0/800	1/800
78	0/800	1/800	0/800
81	0/800	1/800	0/800
85	1/800	0/800	1/800
86	0/800	1/800	0/800
89	1/800	0/800	0/800
96	1/800	0/800	0/800

3. Atribut Kebangasaan

Tabel 3.10. Atribut Kebangasaan

Kebangsaan	Released	Isolated	Deceased
1(Lokal)	75/800	699/800	15/800
2(Asing)	2/800	8/800	1/800

3. Contoh Kasus Perhitungan *Naïve Bayes*

Untuk memudahkan dalam pemahaman perhitungan *Naïve Bayes* secara manual dibuat studi kasus dari Tabel 4.9:

Tabel 3.11. Contoh Pasien

umur	jk	kebangsaan	status
25	Perempuan	Lokal	?

Berikut adalah tahapan untuk menghitung data di atas menggunakan *naïve bayes*

1. Tahap pertama menghitung jumlah *class/label* (probabilitas awal)
- $P(\text{status_pasien} / \text{jumlah data})$ merupakan perhitungan jumlah dari *released*, *isolated*, dan *deceased* dari tabel data *training*. Dari perhitungan di atas maka didapatkan hasil sebagai berikut:

$$P(\text{Released}) = 77/800 = 0.09625$$

$$P(\text{Isolated}) = 707/800 = 0.88375$$

$$P(\text{Deceased}) = 16/800 = 0.0200$$

2. Tahap kedua menghitung jumlah *record atribut*

- P (Jenis Kelamin “perempuan” yang terindikasi status pasien “*released, isolated, dan deceased*”/jumlah total data)

$$P(\text{Jk} = \text{"Perempuan"} | \text{Released}) = 27/800 = 0.0338$$

$$P(\text{Jk} = \text{"Perempuan"} | \text{Isolated}) = 92/800 = 0.1150$$

$$P(\text{Jk} = \text{"Perempuan"} | \text{Deceased}) = 3/800 = 0.00375$$

(Perhitungan di atas adalah jumlah dari data jenis kelamin “perempuan” dengan status pasien “*released, isolated, dan deceased*” dibagi dengan jumlah total data)

- P (Umur “25” yang terindikasi status pasien “*released, isolated, dan deceased*” /Jumlah total data)

$$P(\text{Umur} = \text{"25"} | \text{Released}) = 1/800 = 0.00125$$

$$P(\text{Umur} = \text{"25"} | \text{Isolated}) = 10/800 = 0.0125$$

$$P(\text{Umur} = \text{"25"} | \text{Deceased}) = 0/800 = 0.0000$$

(Perhitungan di atas adalah jumlah dari data umur “25” dengan status pasien “*released, isolated, dan deceased*” dibagi dengan jumlah total data)

- P (Kebangsaan “Lokal” yang terindikasi status pasien “*released, isolated, dan deceased*” /Jumlah total data)

$$P(\text{Kebangsaan} = \text{"Lokal"} | \text{Released}) = 75/800 = 0.0938$$

$$P(\text{Kebangsaan} = \text{"Lokal"} | \text{Isolated}) = 699/800 = 0.8738$$

$$P(\text{Kebangsaan} = \text{"Lokal"} | \text{Deceased}) = 15/800 = 0.0188$$

(Perhitungan di atas adalah jumlah dari data kebangsaan “Lokal” dengan status pasien “*released, isolated, dan deceased*” dibagi dengan jumlah total data)

3. Tahap ketiga adalah mengkalikan semua jumlah data *record* dari perempuan, umur, dan kebangsaan

- $P(X | \text{Released}) = 0.0338 \times 0.00125 \times 0.0938 = 0.00000396305$

(Perhitungan di atas adalah jumlah dari *record* yang didapatkan pada tahap dua yaitu perempuan (*released*) x umur (*released*) x kebangsaan(*released*)

- $P(X|Isolated) = 0.1150 \times 0.0125 \times 0.8738 = 0.0012560875$

(Perhitungan di atas adalah jumlah dari *record* yang didapatkan pada tahap dua yaitu perempuan (*isolated*) x umur (*isolated*) x kebangsaan (*isolated*)

- $P(X|Deceased) = 0.00375 \times 0 \times 0.0188 = 0$

(Perhitungan di atas adalah jumlah dari *record* yang didapatkan pada tahap dua yaitu perempuan (*deceased*) x umur (*deceased*) x kebangsaan (*deceased*)

4. Tahap keempat adalah mengkalikan hasil dari perhitungan semua jumlah data *record* dengan probabilitas awal pada tahap satu

- $P(\text{Semua jumlah data } record) * P(\text{Probabilitas awal})$

$$P(X|Released) * P(Released) = 0.00000396305 \times 0.0963 = 0.000000381641715$$

(Perhitungan di atas adalah hasil dari perhitungan *released* yang didapatkan pada tahap tiga kemudian dikalikan dengan probabilitas awal pada tahap satu)

- $P(X|Isolated) * P(Isolated) = 0.0012560875 \times 0.8838 = 0.0011101301325$

(Perhitungan di atas adalah hasil dari perhitungan *isolated* yang didapatkan pada tahap tiga kemudian dikalikan dengan probabilitas awal pada tahap satu)

- $P(X|Deceased) * P(Deceased) = 0 \times 0.0200 = 0$

(Perhitungan di atas adalah hasil dari perhitungan *deceased* yang didapatkan pada tahap tiga kemudian dikalikan dengan probabilitas awal pada tahap satu)

Tabel 3.12. Tabel Ranking

Rank	Probabilitas	Klasifikasi
1	0.0011101301325	Isolated (1)
2	0.000000381641715	Released (0)
3	0	Deceased (2)

Berdasarkan Tabel 3.12, untuk pasien dengan data di Tabel 3.11 diprediksi memiliki *current_status isolated* (1) karena pasien dengan kategori *isolated* (1) memiliki hasil probabilitas lebih banyak dibanding *released* (1) dan *deceased* (2).

3.3.6.2 Perhitungan K-Nearest Neighbor

Untuk memudahkan dalam pemahaman perhitungan *K-Nearest Neighbor* secara manual dibuat tabel dan studi kasus sebagai berikut:

Tabel 3.13. Contoh Kasus Perhitungan *K-Nearest Neighbor*

<i>no patient</i>	<i>age</i>	<i>gender</i>	<i>nationality</i>	<i>current_status</i>
1	86	1	2	1
2	68	2	1	1
3	43	2	1	1
4	78	1	1	1
5	67	2	1	1
6	66	1	1	2
7	20	2	1	0
8	45	1	1	0
9	24	1	1	0
10	69	1	2	0
11	70	2	2	0
12	45	2	1	0
13	16	1	1	0
14	27	1	1	1
15	59	1	1	0
16	65	1	1	1
17	75	1	1	1
18	45	1	1	0
19	24	1	1	0
20	54	1	1	?

Dari Tabel 3.13 diketahui bahwa *current_status* pada nomor pasien ke 20 tidak dapat terdefinisi, maka dari itu dari 19 data training di atas dilakukan prediksi *current_status* nomor pasien 20 dengan menggunakan *K-Nearest Neighbor*. Perhatikan perhitungan *distance* pada tabel di bawah:

Tabel 3.14. Perhitungan Distance *K-Nearest Neighbor*

<i>no patient</i>	<i>age</i>	<i>gender</i>	<i>natio nality</i>	<i>current_ status</i>	<i>distance</i>
1	86	1	2	1	$\sqrt{(86 - 54)^2 + (1 - 1)^2 + (2 - 1)^2} = 32.02$
2	68	2	1	1	$\sqrt{(68 - 54)^2 + (2 - 1)^2 + (1 - 1)^2} = 14.04$
3	43	2	1	1	$\sqrt{(43 - 54)^2 + (2 - 1)^2 + (1 - 1)^2} = 11.05$
4	78	1	1	1	$\sqrt{(78 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 24.00$
5	67	2	1	1	$\sqrt{(67 - 54)^2 + (2 - 1)^2 + (1 - 1)^2} = 13.04$
6	66	1	1	2	$\sqrt{(66 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 12.00$
7	20	2	1	0	$\sqrt{(20 - 54)^2 + (2 - 1)^2 + (1 - 1)^2} = 34.01$
8	45	1	1	0	$\sqrt{(45 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 9.00$
9	24	1	1	0	$\sqrt{(24 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 30.00$
10	69	1	2	0	$\sqrt{(69 - 54)^2 + (1 - 1)^2 + (2 - 1)^2} = 15.03$
11	70	2	2	0	$\sqrt{(70 - 54)^2 + (2 - 1)^2 + (2 - 1)^2} = 16.06$
12	45	2	1	0	$\sqrt{(45 - 54)^2 + (2 - 1)^2 + (1 - 1)^2} = 9.06$
13	16	1	1	0	$\sqrt{(16 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 38.00$
14	27	1	1	1	$\sqrt{(27 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 27.00$
15	59	1	1	0	$\sqrt{(59 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 5.00$
16	65	1	1	1	$\sqrt{(65 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 11.00$
17	75	1	1	1	$\sqrt{(75 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 21.00$
18	45	1	1	0	$\sqrt{(45 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 9.00$
19	24	1	1	0	$\sqrt{(24 - 54)^2 + (1 - 1)^2 + (1 - 1)^2} = 30.00$

Dari hasil perhitungan *distance* pada Tabel 3.14, kemudian diurutkan dari *distance* terkecil ke terbesar yang diikuti dengan *current_status*. Berikut tabel urutan dari *distance* terkecil ke yang terbesar:

Tabel 3.15. Urutan Distance Perhitungan *K-Nearest Neighbor*

<i>current_status</i>	<i>distance</i>
0	5.00
0	9.00
0	9.00

<i>current_status</i>	<i>distance</i>
0	9.06
1	11.00
1	11.05
2	12.00
1	13.04
1	14.04
0	15.03
0	16.06
1	21.00
1	24.00
1	27.00
0	30.00
0	30.00
1	32.02
0	34.01
0	38.00

Dari tabel hasil urutan *distance* di atas kemudian tentukan nilai K, K merupakan tetangga terdekat dari data yang diprediksi, disini penulis menentukan nilai K=1, K=3, K=5, K=7, K=9, K=11, K=13, K=15. Berikut tabel dengan nilai K:

Tabel 3.16. Prediksi Nilai K Dari *K-Nearest Neighbor*

<i>current_status</i>	<i>distance</i>	k1	k3	k5	K7	K9	K11	K13	K15
0	5.00	0	0	0	0	0	0	0	0
0	9.00		0	0	0	0	0	0	0
0	9.00		0	0	0	0	0	0	0
0	9.06			0	0	0	0	0	0
1	11.00			1	1	1	1	1	1
1	11.05				1	1	1	1	1
2	12.00				2	2	2	2	2
1	13.04					1	1	1	1
1	14.04					1	1	1	1
0	15.03						0	0	0
0	16.06						0	0	0
1	21.00							1	1
1	24.00							1	1
1	27.00								1
0	30.00								0

<i>current_status</i>	<i>distance</i>	k1	k3	k5	K7	K9	K11	K13	K15
0	30.00								
1	32.02								
0	34.01								
0	38.00								

Dari Tabel 3.17 dapat disimpulkan pasien dengan data:

Tabel 3.17. Contoh Soal Dari *K-Nearest Neighbor*

<i>no patient</i>	<i>age</i>	<i>gender</i>	<i>nationality</i>	<i>current_status</i>
20	54	1	1	?

Jika $k=1$ diprediksi *current_status* dari pasien adalah *released* (0) karena *released* (0) lebih dominan dibanding *isolated* (1) dan *deceased* (2).

Jika $k=3$ diprediksi *current_status* dari pasien adalah *released* (0) karena *released* (0) lebih dominan dibanding *isolated* (1) dan *deceased* (2).

Jika $k=5$ diprediksi *current_status* dari pasien adalah *released* (0) karena *released* (0) lebih dominan dibanding *isolated* (1) dan *deceased* (2).

Jika $k=7$ diprediksi *current_status* dari pasien adalah *released* (0) karena *released* (0) lebih dominan dibanding *isolated* (1) dan *deceased* (2).

Jika $k=9$ diprediksi *current_status* dari pasien adalah *released* (0) atau *isolated* (1) karena angka 0 dan 1 mempunyai jumlah yang sama.

Jika $k=11$ diprediksi *current_status* dari pasien adalah *released* (0) karena *released* (0) lebih dominan dibanding *isolated* (1) dan *deceased* (2).

Jika $k=13$ diprediksi *current_status* dari pasien adalah *released* (0) atau *isolated* (1) karena angka 0 dan 1 mempunyai jumlah yang sama.

Jika $k=15$ diprediksi *current_status* dari pasien adalah *released* (0) atau *isolated* (1) karena angka 0 dan 1 mempunyai jumlah yang sama.