

Prediksi Tingkat Kesembuhan Pasien Covid-19 Berdasarkan Riwayat Vaksin Menggunakan Metode Naïve Bayes

Candra Gudiato¹, Sri Yulianto Joko Prasetyo², Hindriyanto Dwi Purnomo³

Magister Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia
 Email: ¹972020001@student.uksw.edu, ²sri.yulianto@uksw.edu, ³hindriyanto.purnomo@uksw.edu
 Email Penulis Korespondensi: ¹972020001@student.uksw.edu
 Submitted: 24/06/2022; Accepted: 30/06/2022; Published: 30/06/2022

Abstrak—Covid-19 telah menghebohkan dunia sejak kemunculannya pertama kali pada akhir Desember 2019. Memasuki tahun 2022, masyarakat *global* sudah lebih siap menghadapi pandemi covid-19, terlebih dengan adanya program vaksinasi *massal* pada negara-negara di Dunia, tak terkecuali di Indonesia. Yang menjadi persoalan selanjutnya adalah seberapa efektif vaksin dalam menghadapi virus covid-19. Parameter utama yang digunakan adalah melihat tingkat kesembuhan pasien yang terkena covid-19 berdasarkan riwayat dosis vaksin yang telah diterima pasien tersebut. Pada penelitian ini menggunakan teknik *data mining*, yaitu menggunakan algoritma *Naïve Bayes*. Hasil pengujian menunjukkan tingkat akurasi dari algoritma Naïve Bayes adalah 98.14%. Adapun hasil prediksi menunjukkan bahwa tingkat kesembuhan pasien yang sudah menerima vaksin, baik dosis 1, dosis 2, maupun dosis 3 (*booster*) lebih tinggi dibandingkan mereka yang belum divaksin sama sekali (dosis 0). Hasil penelitian ini diharapkan dapat memberi gambaran kepada masyarakat dan pemerintah akan manfaat vaksinasi dalam menghadapi virus Covid-19.

Kata Kunci: Covid-19; Vaksin; Data Mining; Naïve Bayes; Prediksi

Abstract—Covid-19 has shocked the world since it first appeared at the end of December 2019. At the beginning of 2022, the global community is more prepared to face the COVID-19 pandemic, especially with the mass vaccination program in countries around the world, including Indonesia. The next issue is how effective the vaccine is in dealing with the COVID-19 virus. The main parameter used is to see the recovery rate of patients affected by COVID-19 based on the history of vaccine doses that have been received by the patient. In this study using data mining techniques, namely using the Naïve Bayes algorithm. The test results show the accuracy of the Naïve Bayes algorithm is 98.14%. The prediction results show that the recovery rate of patients who have received the vaccine, either dose 1, dose 2, or dose 3 (booster) is higher than those who have not been vaccinated at all (dose 0). The results of this study are expected to provide an overview to the public and the government about the benefits of vaccination in dealing with the Covid-19 virus.

Keywords: Covid-19; Vaccine; Data Mining; Naïve Bayes; Prediction

1. PENDAHULUAN

Akhir Desember 2019, dilaporkan kasus *pneumonia* yang belum pernah terjadi sebelumnya di Wuhan, Provinsi Hubei, China. Sumber dari penularan ini tidak diketahui secara pasti, namun kasus pertama dikaitkan dengan pasar ikan di Wuhan. Setelah kejadian tersebut, kasus serupa kemudian meningkat dengan pesat dan mulai menyebar di berbagai provinsi di China dan negara-negara tetangga, seperti Jepang, Thailand, dan Korea Selatan.

Awalnya, virus ini dinamakan sebagai 2019 *novel Coronavirus* (2019-nCov), kemudian WHO mengumumkan nama baru yaitu *Coronavirus Disease 2019* (COVID-19) yang disebabkan oleh virus *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-Cov-2)[1]. Jumlah kasus yang diakibatkan oleh Covid-19 kemudian terus meningkat dengan pesat sampai kepada tahap pandemi global. Berdasarkan data yang dikeluarkan oleh situs covid-19 seperti yang ditunjukkan pada gambar 1, khususnya wilayah Indonesia pada 30 Maret 2022 terdapat 6.009.486 kasus yang terkonfirmasi positif covid-19, dengan 5.742.931 di antaranya telah sembuh, dan jumlah pasien meninggal sebanyak 155.000 jiwa[2].



Gambar 1. Situasi Penyebaran Virus Covid-19 di Indonesia (Sumber: Satgas Covid-19 Indonesia)

Untuk menghadapi pandemi Covid-19, para peneliti dari seluruh dunia melakukan segala daya upaya untuk menghasilkan vaksin yang dapat melawan virus tersebut. Vaksinasi diprioritaskan untuk kelompok yang tergolong rentan. Di saat yang sama, lebih dari 200 kandidat vaksin telah diproduksi dengan 60 di antaranya telah masuk dalam tahap uji klinis[3]. Saat ini vaksinasi massal telah dilakukan pada negara-negara di dunia terlepas dari pro kontra dari masyarakatnya. Berdasarkan data yang dihimpun oleh Google news per 30 Maret 2022 yang dapat dilihat pada gambar 2, jumlah populasi yang telah mendapatkan vaksin secara lengkap 2 dosis (*Fully vaccinated*) sebanyak kurang lebih

4,542 miliar jiwa (58,4% dari total penduduk dunia). Khusus untuk wilayah Indonesia, populasi yang telah mendapatkan vaksin lengkap sekitar 159 juta jiwa (59,5% dari total penduduk Indonesia)[4].



Gambar 2. Situasi Vaksinasi di Dunia (Sumber: Google News)

Tak hanya penelitian di bidang medis saja yang telah dilakukan, ada pula penelitian non-medis yaitu dari kalangan akademisi yang mengupas hal-hal yang terkait dengan pandemi Covid-19 dari berbagai aspek, seperti ekonomi, sosial, budaya, teknologi, dan lain sebagainya. Salah satu penelitian terdahulu mengenai Covid-19 yang juga menjadi acuan dasar dari penelitian ini adalah penelitian yang dilakukan oleh Naim Rochmawati, dkk. yang dipublikasikan pada tahun 2020 berjudul “Covid Symptom Severity Using Decision Tree”. Pada penelitian tersebut, mereka melakukan klasifikasi tingkat keparahan (*symptom severity*) yang dialami penderita covid-19 dengan menggunakan *Decision Tree*. Algoritma dari *Decision Tree* yang digunakan adalah J48 dan *Hoeffding Tree*. Adapun *dataset* penelitian diambil dari situs Kaggle dengan pengolahan menggunakan *tools* aplikasi Weka. Hasil *output* penelitian berupa kesimpulan apakah seorang pasien covid-19 termasuk dalam kategori tingkat keparahan ringan (*mild*), sedang (*moderate*), berat (*severe*), atau tidak terkena covid (*not-covid*). Selain itu, hasil penelitian yang mereka lakukan juga menunjukkan bahwa *Hoeffding Tree* memiliki *accuracy*, *precision*, dan *recall* yang lebih baik dibandingkan J48[5].

Kondisi penelitian pada tahun 2020 lebih cenderung berfokus pada virusnya, yaitu efek maupun gejala dari seorang pasien yang terkena Covid-19. Pada tahun 2021, vaksinasi massal mulai dilakukan pada negara-negara yang ada di dunia, tak terkecuali di Indonesia. Maka dari itu, arah penelitian pelan-pelan mulai bergeser fokus pada vaksinnya, bukan pada virusnya lagi. Bahkan memasuki tahun 2022 sudah dilakukan Vaksinasi tahap ke-3 atau lebih familiar dikenal dengan nama vaksin *Booster*. Salah satu pertanyaan dasar yang cukup sering ditanyakan terkait vaksin adalah seberapa efektif vaksinasi untuk menghadapi virus Covid-19, terutama perbandingan seseorang yang hanya menerima vaksinasi tahap pertama, tahap kedua, maupun yang sudah mendapat *booster*. Apakah dengan vaksinasi tingkat kesembuhan para penderita covid-19 dapat lebih meningkat.

Atas kondisi yang diuraikan pada penjelasan di atas, maka akan dilakukan penelitian untuk memprediksi tingkat kesembuhan pasien Covid-19 berdasarkan riwayat vaksin yang telah diterima menggunakan metode *data mining*, yaitu *Naïve Bayes*. *Data mining* adalah metode dari kumpulan prosedur untuk menemukan pola (*pattern*) dari sejumlah besar data yang awalnya tidak diketahui atau tersembunyi menjadi suatu informasi yang berpotensi berguna sehingga dapat digunakan untuk membantu dalam menentukan keputusan strategis[6][7]. *Naïve Bayes* termasuk dalam *data mining* teknik *classification*. *Classification* memungkinkan mempelajari pola-pola dari data historis, seperti ciri-ciri, variabel-variabel, fitur-fitur, pada berbagai karakteristik *item-item* yang sudah diberi label sebelumnya dengan tujuan untuk menempatkan objek-objek baru yang belum teridentifikasi ke dalam kelas atau kelompok masing-masing[8].

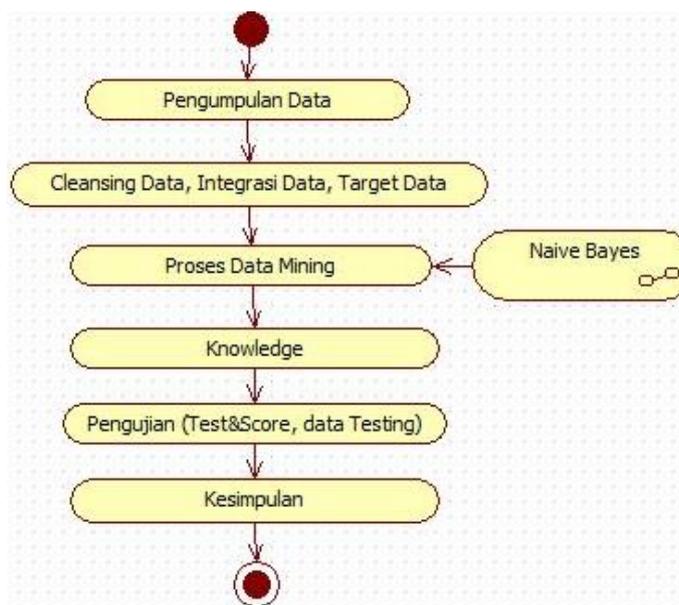
Naïve Bayes dapat digunakan untuk melakukan prediksi atas suatu kejadian di masa mendatang, yang dibuktikan pada penelitian yang dilakukan oleh Fajar Edi Prabowo dan Achmad Kodar yang dipublikasikan pada tahun 2019 berjudul “Analisis Prediksi Masa Studi Mahasiswa Menggunakan Algoritma *Naïve Bayes*”. Penelitian tersebut menggunakan bantuan *tools* aplikasi *Rapid Miner* dalam melakukan prediksi penentuan probabilitas seorang mahasiswa apakah lulus “Cepat”, “Tepat Waktu”, atau “Terlambat”, dengan tingkat akurasi rata-rata sebesar 82,26%[9].

Prediksi menggunakan *Naïve Bayes* juga diterapkan pada penelitian yang dilakukan oleh Mohammad Guntur, dkk. pada publikasi yang berjudul “Prediksi Harga Emas dengan Menggunakan Metode *Naïve Bayes* dalam Investasi untuk Meminimalisasi Resiko”. Penelitian tersebut menggunakan bantuan *tools* aplikasi *Rapid Miner* untuk memprediksi harga emas di masa mendatang sehingga pengambil keputusan dapat menentukan apakah harus membeli atau menjual emas dengan tingkat akurasi sebesar 75%[10].

Penelitian ini akan memanfaatkan informasi yang telah dihimpun berupa *dataset* untuk kemudian dilakukan proses *data mining* menggunakan metode *Naïve Bayes*. Hasil keluaran dari penelitian ini berupa prediksi kesembuhan pasien berdasarkan riwayat vaksin yang diperoleh (belum divaksin, dosis 1, dosis 2, *booster*). Dari hasil penelitian, diharapkan masyarakat akan lebih sadar akan manfaat dari vaksinasi dan pemerintah lebih pro-aktif dalam memfasilitasi pemberian vaksinasi.

2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan melalui beberapa tahap, yaitu pengumpulan data, pengolahan dan analisis data (*cleansing*, *integrasi data*, *target data*, proses *data mining*, *knowledge*), dan evaluasi (Pengujian *Test and Score* dan Pengujian menggunakan *Data Testing*). Adapun penerapan metode *Naïve Bayes* bekerja dalam tahapan proses *data mining*. Tahap-tahap pada penelitian digambarkan pada gambar 3 di bawah ini:



Gambar 3. Tahap - Tahap Penelitian

2.1 Pengumpulan Data

Sumber data yang digunakan adalah data pasien Covid-19 dan data vaksinasi di wilayah Kabupaten Bengkayang dengan rentang waktu bulan Oktober 2021 – Maret 2022 yang diperoleh dari Dinas Kesehatan dan Keluarga Berencana Kabupaten Bengkayang - Kalimantan Barat. Sumber data inilah yang akan menjadi *data training* dalam proses *data mining*.

2.2 Cleansing Data

Data yang tersedia selanjutnya akan dilakukan tahap *cleansing data* atau pembersihan data sehingga data yang nantinya akan diolah benar-benar relevan dengan yang dibutuhkan. Biasanya *cleansing data* memperbaiki & membuang data yang tidak lengkap (*incomplete*), menyimpang (*noise*), dan tidak cocok (*inconsistent*).

2.3 Integrasi Data

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru[11]. Pada tahap ini, sumber data yang terpisah-pisah pada beberapa dokumen Microsoft Excel akan digabungkan menjadi sebuah dokumen baru.

2.4 Target Data

Tahap ini menentukan data yang siap digunakan dalam proses *data mining*, dimana *target data* menjadi *output* yang diinginkan.

2.5 Proses data Mining

Tahap ini merupakan tahap pembentukan alur dari *Data Mining* yang dimana akan menggunakan algoritma *Naïve Bayes* menggunakan *tools Orange*. Algoritma *Naïve Bayes* dipilih karena sumber data yang berhasil dihimpun cukup besar, dimana ciri khas dari *Naïve Bayes* adalah semakin banyak data, maka akan semakin akurat.

Naïve Bayes merupakan teori yang dikemukakan oleh Thomas Bayes tahun 1950 yang merupakan kondisi probabilitas suatu kejadian hipotesis bergantung pada kejadian lain sebagai bukti. Pada dasarnya, teori tersebut



menyatakan bahwa kejadian di masa depan dapat diprediksi dengan syarat kejadian sebelumnya telah terjadi. Teori Naïve Bayes memiliki kemampuan klasifikasi yang serupa dengan *Decision Tree* dan *Neural Network* bahkan algoritma Naïve Bayes memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar[12]. Secara umum teori Naïve Bayes ditulis dalam bentuk:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Keterangan:

- H adalah sebuah hipotesa dari data yang berupa suatu *class* spesifik.
- X adalah *class* yang datanya belum diketahui.
- P(H|X) yaitu kemungkinan hipotesa H berdasarkan kondisi X (*posteriori probabilitas*)
- P(X|H) yaitu probabilitas X berdasarkan kondisi pada hipotesis H (Rumus *Likelihood*)
- P(H) yaitu probabilitas hipotesis H (*prior probabilitas*)
- P(X) yaitu probabilitas dari *evidence* X [13].

Adapun *dataset* yang digunakan sebagai data latih (*training data*) bersumber dari Dinas Kesehatan dan Keluarga Berencana Kabupaten Bengkayang Provinsi Kalimantan Barat. Pengolahan menggunakan *tools Orange*. *Orange* adalah sebuah aplikasi untuk *machine learning* dan *data mining*. *Orange* dikembangkan tahun 1997 oleh Janez Demsar dan Blaz Zupan. *Orange* dikembangkan kembali dengan menggunakan bahasa pemrograman *Python* setelah sebelumnya dikembangkan dengan bahasa C++. *Python* yang merupakan bahasa pemrograman yang modern dipilih karena mempunyai sintaks yang sederhana yang membuatnya mudah untuk dipelajari. Pembuatan program dengan *Python* dapat dilakukan dengan cepat, sehingga sangat cocok dalam pembuatan metode baru, dan relatif mudah mengembangkan *Python* dengan modul yang ditulis dengan bahasa C atau C++[14].

Hasil analisis diuji & dievaluasi menggunakan *Confusion Matrix*. *Confusion Matrix* biasanya digunakan untuk evaluasi kinerja model klasifikasi yang berdasarkan pada kemampuan akurasi prediktif suatu model. Akurasi dinyatakan dalam persentase, sehingga akurasi 100% artinya semua kasus yang tercakup oleh aturan klasifikasi, berhasil diklasifikasikan dengan benar ke dalam kelas yang diprediksinya. Untuk mendapatkan nilai akurasi prediktif diperlukan perhitungan jumlah kasus yang diprediksikan dengan benar dan jumlah kasus yang diprediksikan dengan salah. Perhitungan tersebut ditabulasikan ke dalam tabel yang disebut *confusion matrix* seperti tabel di bawah ini[15]:

Tabel 1. Confusion Matrix

Actual Class	Predicted Class	
	YES	NO
YES	True Positive (A)	False Negative (B)
NO	False Positive (C)	True Negative (D)

Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$Akurasi = (A+D)/(A+B+C+D) \tag{2}$$

Precision didefinisikan sebagai rasio *item* relevan yang dipilih terhadap semua *item* yang terpilih. *Precision* dapat diartikan sebagai kecocokan antara permintaan informasi dengan jawaban terhadap permintaan tersebut. Rumus *precision* adalah:

$$Precision = A/(C + A) \tag{3}$$

Recall didefinisikan sebagai rasio dari *item* relevan yang dipilih terhadap total jumlah *item* relevan yang tersedia. *Recall* dihitung dengan rumus:

$$Recall = A/(A + D) \tag{4}$$

Precision dan *Recall* dapat diberi nilai dalam bentuk angka dengan menggunakan perhitungan persentase (0-100%) atau dengan menggunakan bilangan antara 0-1. Suatu analisis akan dianggap baik jika memiliki nilai *precision* dan *recall*-nya tinggi[16].

2.6 Knowledge

Setelah melalui setiap tahapan maka akan dihasilkan sebuah pengetahuan baru (*Knowledge*) yang selanjutnya dapat dimanfaatkan untuk kepentingan penelitian.

2.7 Pengujian dan Evaluasi Algoritma (Test and Score)

Knowledge yang telah dibentuk akan diuji untuk menentukan tingkat akurasi dari pengetahuan baru tersebut. Pengujian dilakukan menggunakan *K-Fold Cross Validation*. *K-Fold Cross Validation* adalah suatu metode validasi silang yang tersedia dalam literatur untuk memilih sampel sebagai *training dataset*. *K-Fold Cross Validation* membagi sampel aktual menjadi K sub sampel yang berukuran sama. Setiap sub sampel diambil sebagai data validasi untuk menguji model klasifikasi dan mengulang proses K. Nilai K adalah parameter tidak tetap yang dipilih oleh pengujian (*user*)[17]. Hasil dari pengujian berupa evaluasi yang ditampilkan dalam bentuk *Confusion Matrix*.

2.8 Pengujian dan Evaluasi menggunakan *Data Testing*

Pengujian dengan *data testing* dilakukan dengan *widjet predictions*, dimana *knowledge* yang telah dihasilkan akan mencoba melakukan prediksi terhadap sekumpulan *data testing* yang telah dipersiapkan.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pada penelitian ini, data yang tersedia sebagai *dataset* berjumlah 2561 *records* dan 6 atribut sebagai *feature* dan 1 atribut sebagai *target data*. Adapun atribut-atribut yang tersedia dapat dilihat pada tabel 2 di bawah ini:

Tabel 2. Atribut Sumber Data

ATRIBUT	DESKRIPSI
No	Penomoran
Jenis Kelamin	Laki-laki atau Perempuan
Usia	Usia Pasien
Gejala	Ringan, Sedang, dan Berat
Jenis Pemeriksaan	Swab PCR atau RDT Antigen
Dosis	Riwayat vaksinasi dosis 0, 1, 2, atau 3
Status Akhir Pasien	Sembuh atau Meninggal

Untuk atribut gejala, penggolongan seseorang bergejala ringan apabila seorang pasien mengalami sakit kepala dan lemas, maupun yang tidak memiliki gejala sama sekali. Penggolongan gejala sedang apabila seorang pasien mengalami sakit tenggorokan, batuk, hilang kemampuan indera penciuman, dan hilang kemampuan indera pengecap. Penggolongan gejala berat apabila seorang pasien mengalami batuk dengan intensitas sering, demam dan sesak napas. Adapun *dataset* yang disediakan dalam bentuk Microsoft Excel seperti pada gambar 4 di bawah ini:

	A	B	C	D	E	F	G
1	No	Jenis Kelamin	Usia	Gejala	Jenis Pemeriksaan	Dosis	Status Akhir Pasien
2	1	Perempuan	13	Sedang	Swab PCR	2	SEMBUH
3	2	Laki-Laki	15	Ringan	Swab PCR	1	SEMBUH
4	3	Perempuan	17	Ringan	Swab PCR	2	SEMBUH
5	4	Perempuan	53	Sedang	RDT Antigen	2	SEMBUH
6	5	Perempuan	20	Ringan	Swab PCR	1	SEMBUH
7	6	Perempuan	1	Sedang	Swab PCR	2	SEMBUH
8	7	Perempuan	2	Berat	Swab PCR	2	SEMBUH
9	8	Perempuan	2	Berat	Swab PCR	2	SEMBUH
10	9	Perempuan	17	Sedang	Swab PCR	2	SEMBUH
11	10	Perempuan	18	Sedang	Swab PCR	1	SEMBUH
12	11	Perempuan	28	Berat	RDT Antigen	1	SEMBUH
13	12	Perempuan	31	Sedang	Swab PCR	2	SEMBUH
14	13	Perempuan	21	Ringan	Swab PCR	1	SEMBUH
15	14	Perempuan	43	Berat	RDT Antigen	2	SEMBUH
16	15	Perempuan	14	Ringan	RDT Antigen	1	SEMBUH
17	16	Perempuan	25	Ringan	Swab PCR	3	SEMBUH
18	17	Perempuan	54	Sedang	Swab PCR	3	SEMBUH
19	18	Laki-Laki	25	Berat	RDT Antigen	1	SEMBUH
20	19	Perempuan	25	Sedang	Swab PCR	2	SEMBUH
21	20	Perempuan	19	Ringan	RDT Antigen	3	SEMBUH
22	21	Perempuan	19	Berat	RDT Antigen	1	SEMBUH
23	22	Perempuan	71	Ringan	Swab PCR	1	SEMBUH

Gambar 4. *Dataset* Penelitian

3.2 Cleansing Data

Pada tahap ini dilakukan perbaikan beberapa *typo* yang muncul, karena *Orange* bersifat *case sensitive*. Data yang diberikan lengkap dan tidak terdapat *record* yang isinya *missing*, sehingga 2561 data yang tersedia akan dipakai semua pada tahap selanjutnya.

3.3 Integrasi Data

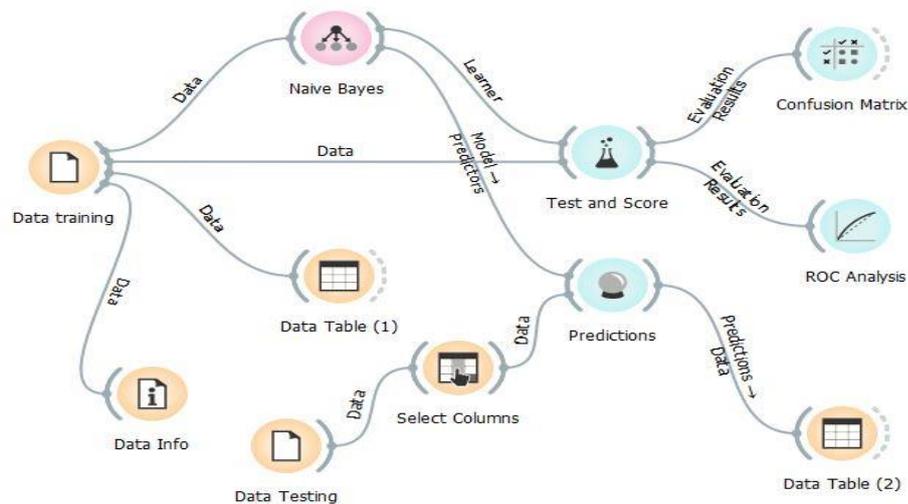
Data yang diberikan terpisah menjadi dua dokumen Microsoft Excel, yaitu antara dokumen pasien Covid-19 dan dokumen Vaksinasi. Integrasi dilakukan dengan menggabungkan kedua dokumen dengan menyamakan NIK.

3.4 Target Data

Target data yang dipilih adalah atribut “Status Akhir Pasien” yang bernilai “Sembuh” atau “Meninggal”. Atribut tersebut dipilih sebagai target untuk mengidentifikasi tingkat kesembuhan. Selain itu, atribut “No” dan “Jenis Pemeriksaan” tidak diikutsertakan dalam input yang digunakan karena apabila dimasukkan sebagai parameter akan membuat hasil *mining* menjadi bias.

3.5 Proses data Mining

Proses data mining secara implisit dilakukan oleh *tools* Orange dengan rancangan *Diagram mining* seperti gambar 5 di bawah ini:



Gambar 5. Diagram Mining Naïve Bayes

Adapun proses algoritma *Naïve Bayes* yang bekerja untuk menyelesaikan masalah diuraikan secara singkat melalui *pseudocode* yang terdapat pada gambar 6 di bawah ini:

```

Input   : Dataset (Training Data)
Output  : Status akhir Pasien (Sembuh atau Meninggal)
Step of Naïve Bayes algorithm:
1. Read dataset
2. Calculate probabilitas “perempuan yang sembuh” terhadap semua pasien yang sembuh
3. Calculate probabilitas “perempuan yang meninggal” terhadap semua pasien yang meninggal
4. Repeat step [2] [3] for laki-laki
5. Repeat step [2][3][4] for other (umur, gejala, dosis)
6. Multiplication probabilitas yang mengandung unsur “sembuh” (X)
7. Multiplication probabilitas yang mengandung unsur “meninggal” (Y)
8. Using equation  $P(H|X) = \frac{P(X|H)P(H)}{P(X)}$ 
   Where P(X) is Predictor Prior Probability. If there is no predictor, then can be ignored or P(X)=1
   P(Sembuh|X)= ( P(Perempuan|Sembuh) * P(Laki-laki|Sembuh) * P(Umur 1| Sembuh) * .... * P(Umur
   94|Sembuh) * P( Gejala Ringan|Sembuh) * P(Gejala Sedang|Sembuh) * P(Gejala
   Berat|Sembuh) * P(Dosis 0|Sembuh) * P(Dosis 1|Sembuh) * P(Dosis 2|Sembuh) *
   P(Dosis 3|Sembuh) * P(Sembuh)) / P(X)
   P(Meninggal|Y)= ( P(Perempuan|Meninggal) * P(Laki-laki|Meninggal) * P(Umur 1| Meninggal) * .... *
   P(Umur 94|Meninggal) * P( Gejala Ringan|Meninggal) * P(Gejala Sedang|Meninggal) *
   P(Gejala Berat|Meninggal) * P(Dosis 0|Meninggal) * P(Dosis 1|Meninggal) * P(Dosis
   2|Meninggal) * P(Dosis 3|Meninggal) * P(Meninggal)) / P(Y)
9. If (P(Sembuh|X) > P(Meninggal|Y)
   Then Prediksi Naïve Bayes=“sembuh”;
   Else If (P(meninggal|Y) > P(Sembuh|X)
   Then Prediksi Naïve Bayes= “meninggal”;
   Endif
    
```

Gambar 6. Pseudocode Algoritma Naïve Bayes

3.6 Knowledge

Knowledge yang dihasilkan tidak ditampilkan secara eksplisit, namun aplikasi yang dibangun telah mengingat ke-2561 pola yang dimasukkan sebagai data latih (*training data*) yang akan dapat digunakan untuk memecahkan masalah ataupun memprediksi suatu kejadian dengan *input* variabel yang sama.

3.7 Pengujian dan Evaluasi Algoritma (*Test and Score*)

Pengujian menggunakan *K-Fold Cross Validation* dengan menggunakan nilai K=2, K=3, K=5, K=10, dan K=20. Adapun hasil *Test and Score* disajikan pada tabel 3 di bawah ini:

Tabel 3. Hasil Test & Score

Nilai K	Accuracy	Precision	Recall
2	0.980	0.974	0.980
3	0.981	0.975	0.981
5	0.983	0.980	0.983
10	0.982	0.978	0.982
20	0.981	0.975	0.981
Rata-Rata	0.9814	0.98764	0.9814

Hasil dari *confusion matrix* dinyatakan dalam bentuk Kurva ROC. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC adalah grafik dua dimensi dengan *false positive* sebagai garis horizontal dan *true positive* sebagai garis vertikal. ROC memiliki tingkat nilai akurasi sebagai berikut:

- Nilai Akurasi 0,90 – 1,00 = *excellent classification*
- Nilai Akurasi 0,80 – 0,90 = *good classification*
- Nilai Akurasi 0,70 – 0,80 = *fair classification*
- Nilai Akurasi 0,60 – 0,70 = *poor classification*
- Nilai Akurasi 0,50 – 0,60 = *failure* [16]

Dari hasil yang ditunjukkan pada tabel 3 di atas, dapat diketahui bahwa pada penelitian ini algoritma *Naïve Bayes* memiliki rata-rata akurasi 0,9814 (98,14%) sehingga termasuk dalam kelompok *excellent classification* (akurasi >90%) yang artinya memiliki performa yang sangat baik. Hasil evaluasi dari *confusion matrix* disajikan pada gambar 7 di bawah ini:



Gambar 7. Confusion Matrix Naïve Bayes

Berdasarkan hasil *confusion matrix* pada gambar di atas, terlihat bahwa *Naïve Bayes* melakukan kesalahan prediksi sebanyak 48 records, 42 di antaranya bernilai *False Negative* dan 6 bernilai *false positive* dari total 2561 records. Prediksi secara tepat sebanyak 2513 records dengan 7 records di antaranya bernilai *True Positive* dan 2506 bernilai *True Negative*. Maka apabila dimasukkan ke dalam persamaan pada rumus akurasi yang dijabarkan di atas, maka didapatkan akurasi sebagai berikut:

$$Akurasi = (A+D)/(A+B+C+D)$$

$$Akurasi = (7+2506)/(7+42+6+2506)$$

$$Akurasi = 2513 / 2561$$

$$Akurasi = 0,981257 (98,13\%)$$

Hasil perhitungan menggunakan rumus akurasi yaitu 98,13% memiliki hasil yang kurang lebih sama dengan hasil *test and score* yang ditunjukkan pada tabel 3 di atas sebesar 98,14%.

3.8 Pengujian dan Evaluasi menggunakan *Data Testing*

Pengujian menggunakan *data testing* dilakukan dengan *widget data sampler* sebanyak 30% dari *data training* yang tersedia. Hasil dari pengujian tersebut disajikan pada gambar 8 dan gambar 9 di bawah ini:



	Status Akhir Pasien	Naive Bayes	Naive Bayes (MENINGGAL)	Naive Bayes (SEMBUH)	Jenis Kelamin	Usia	Gejala	Dosis
836	SEMBUH	SEMBUH	0.000467648	0.999532	Perempuan	12	Ringan	2
211	SEMBUH	SEMBUH	0.000482166	0.999518	Laki-Laki	12	Ringan	2
432	SEMBUH	SEMBUH	0.00063171	0.999368	Laki-Laki	10	Ringan	2
1141	SEMBUH	SEMBUH	0.0006422	0.999358	Perempuan	14	Ringan	2
1075	SEMBUH	SEMBUH	0.0006422	0.999358	Perempuan	11	Ringan	2
449	SEMBUH	SEMBUH	0.0006422	0.999358	Perempuan	14	Ringan	2
346	SEMBUH	SEMBUH	0.0006422	0.999358	Perempuan	11	Ringan	2
4	SEMBUH	SEMBUH	0.0006422	0.999358	Perempuan	14	Ringan	2
62	SEMBUH	SEMBUH	0.000650027	0.99935	Perempuan	13	Ringan	2
1339	SEMBUH	SEMBUH	0.000662133	0.999338	Laki-Laki	11	Ringan	2
1224	SEMBUH	SEMBUH	0.000662133	0.999338	Laki-Laki	14	Ringan	2
921	SEMBUH	SEMBUH	0.000662133	0.999338	Laki-Laki	11	Ringan	2
800	SEMBUH	SEMBUH	0.000662133	0.999338	Laki-Laki	14	Ringan	2
1635	SEMBUH	SEMBUH	0.000670203	0.99933	Laki-Laki	13	Ringan	2
1078	SEMBUH	SEMBUH	0.000756067	0.999244	Laki-Laki	12	Ringan	3
1458	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
1365	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
1138	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
944	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
760	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
450	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
345	SEMBUH	SEMBUH	0.000778034	0.999222	Perempuan	16	Ringan	2
1325	SEMBUH	SEMBUH	0.000794714	0.999205	Laki-Laki	10	Ringan	1
1131	SEMBUH	SEMBUH	0.000802181	0.999198	Laki-Laki	16	Ringan	2
742	SEMBUH	SEMBUH	0.000802181	0.999198	Laki-Laki	16	Ringan	2
632	SEMBUH	SEMBUH	0.000802181	0.999198	Laki-Laki	16	Ringan	2
1630	SEMBUH	SEMBUH	0.000807909	0.999192	Perempuan	11	Ringan	1
1044	SEMBUH	SEMBUH	0.000807909	0.999192	Perempuan	14	Ringan	1
328	SEMBUH	SEMBUH	0.000832694	0.999167	Perempuan	19	Ringan	2
825	SEMBUH	SEMBUH	0.000858336	0.999141	Laki-Laki	19	Ringan	2

Gambar 8. Hasil Prediksi Naïve Bayes berdasarkan tingkat kesembuhan tertinggi

	Status Akhir Pasien	Naive Bayes	Naive Bayes (MENINGGAL)	Naive Bayes (SEMBUH)	Jenis Kelamin	Usia	Gejala	Dosis
14	MENINGGAL	MENINGGAL	0.893351	0.106649	Perempuan	79	Berat	0
332	MENINGGAL	MENINGGAL	0.893351	0.106649	Perempuan	81	Berat	0
1464	MENINGGAL	MENINGGAL	0.893351	0.106649	Perempuan	79	Berat	0
717	MENINGGAL	MENINGGAL	0.848126	0.151874	Perempuan	94	Berat	0
1027	MENINGGAL	MENINGGAL	0.761778	0.238222	Laki-Laki	81	Sedang	0
1201	MENINGGAL	MENINGGAL	0.742196	0.257804	Laki-Laki	75	Berat	0
1204	SEMBUH	MENINGGAL	0.508312	0.491688	Perempuan	75	Sedang	0
438	MENINGGAL	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
587	SEMBUH	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
815	MENINGGAL	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
1009	MENINGGAL	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
1225	MENINGGAL	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
1578	MENINGGAL	MENINGGAL	0.503807	0.496193	Perempuan	42	Berat	0
428	SEMBUH	SEMBUH	0.411089	0.588911	Perempuan	62	Berat	0
582	MENINGGAL	SEMBUH	0.411089	0.588911	Perempuan	62	Berat	0
455	SEMBUH	SEMBUH	0.365395	0.634605	Laki-Laki	71	Berat	0
1568	SEMBUH	SEMBUH	0.365395	0.634605	Laki-Laki	64	Berat	0
951	SEMBUH	SEMBUH	0.358332	0.641668	Perempuan	64	Berat	0
175	SEMBUH	SEMBUH	0.354158	0.645842	Laki-Laki	43	Berat	0
754	SEMBUH	SEMBUH	0.354158	0.645842	Laki-Laki	43	Berat	0
785	SEMBUH	SEMBUH	0.354158	0.645842	Laki-Laki	43	Berat	0
1151	SEMBUH	SEMBUH	0.354158	0.645842	Laki-Laki	43	Berat	0
305	MENINGGAL	SEMBUH	0.347193	0.652807	Perempuan	43	Berat	0
1367	SEMBUH	SEMBUH	0.347193	0.652807	Perempuan	43	Berat	0
1735	MENINGGAL	SEMBUH	0.347193	0.652807	Perempuan	43	Berat	0
985	SEMBUH	SEMBUH	0.340763	0.659237	Perempuan	77	Sedang	0
1488	SEMBUH	SEMBUH	0.340763	0.659237	Perempuan	78	Sedang	0

Gambar 9. Hasil Prediksi Naïve Bayes berdasarkan tingkat kesembuhan terendah

Berdasarkan gambar 8 dan gambar 9 di atas, kolom “Status Akhir Pasien” adalah kejadian aktual (fakta) yang terjadi di lapangan. Kolom “Naïve Bayes” menunjukkan hasil prediksi berdasarkan *data training* yang tersedia. Sementara itu, untuk kolom “Naïve Bayes (MENINGGAL)” menunjukkan probabilitas (nilai tertinggi 1) seseorang akan meninggal menurut perhitungan Naïve Bayes, sedangkan kolom “Naïve Bayes (SEMBUH)” menunjukkan probabilitas seseorang akan sembuh.

Berdasarkan perbandingan hasil yang ditunjukkan pada kedua gambar di atas, dapat diketahui bahwa pasien yang meninggal didominasi oleh mereka yang belum divaksin (dosis 0). Selain itu probabilitas kesembuhan juga cukup rendah dibandingkan dengan pasien yang sudah menerima vaksin, baik dosis 1, dosis 2, maupun dosis 3 (*booster*).

Pada gambar 9 juga terlihat bahwa terdapat kesalahan prediksi yang dilakukan oleh metode *Naïve Bayes* dengan fakta di lapangan, di mana hasil prediksi adalah pasien sembuh, namun faktanya pasien tersebut meninggal. Hal tersebut terjadi dikarenakan tingkat akurasi *Naïve Bayes* seperti yang ditunjukkan pada tabel 3 di atas adalah 98,14%.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menggunakan data pasien covid-19 dan vaksinasi di wilayah Kabupaten Bengkayang sebanyak 2561 *records* menunjukkan bahwa algoritma *Naïve Bayes* memiliki kinerja akurasi sebesar 0,9814 atau 98,14%. Hasil penelitian juga menunjukkan bahwa tingkat kesembuhan pasien covid-19 semakin meningkat apabila pasien sudah divaksin, dimana dalam penelitian ini pasien yang meninggal semuanya didominasi oleh mereka yang belum divaksin (dosis 0). Selain faktor riwayat vaksin, terdapat faktor lain yang juga mempengaruhi hasil prediksi yaitu jenis kelamin, usia, dan gejala. Dengan hasil penelitian yang telah ditunjukkan, diharapkan masyarakat lebih sadar akan manfaat dari vaksinasi dalam menghadapi virus Covid-19 yang dapat meningkatkan kemungkinan seseorang untuk sembuh. Selain itu pemerintah juga diharapkan lebih pro-aktif dalam memfasilitasi pemberian vaksin, terutama untuk daerah-daerah yang belum terjangkau agar pandemi Covid-19 dapat segera ditanggulangi dan menekan angka kematian. Untuk penelitian dan pengembangan selanjutnya diharapkan dapat menggunakan sumber data yang lebih banyak dengan cakupan wilayah yang lebih luas. Selain itu dapat dipadukan dengan algoritma *data mining* yang lain agar dapat menemukan algoritma yang terbaik dalam menentukan tingkat kesembuhan pasien covid-19.

REFERENCES

- [1] A. Susilo *et al.*, “Coronavirus Disease 2019 : Tinjauan Literatur Terkini,” *J. Penyakit Dalam Indones.*, vol. 7, no. 1, pp. 45–67, 2020.
- [2] Satuan Tugas Penanganan Covid-19, “Situasi virus Covid-19 di Indonesia,” 2022. <https://covid19.go.id/> (accessed Mar. 30, 2022).
- [3] H. Abebe, S. Shitu, and A. Mose, “Understanding of COVID-19 vaccine knowledge, attitude, acceptance, and determinates of COVID-19 vaccine acceptance among adult population in Ethiopia,” *Infect. Drug Resist.*, vol. 14, no. June, pp. 2015–2025, 2021, doi: 10.2147/IDR.S312116.
- [4] Google News, “Coronavirus (Covid-19) Worldwide Statistics,” 2022. <https://news.google.com/covid19/map?hl=en-ID&gl=ID&ceid=ID%3Aen> (accessed Mar. 30, 2022).
- [5] N. Rochmawati *et al.*, “Covid Symptom Severity Using Decision Tree,” 2020, doi: 10.1109/ICVEE50212.2020.9243246.
- [6] Y. Huang, H. Liu, and J. Pan, “Identification of data mining research frontier based on conference papers,” *Int. J. Crowd Sci.*, vol. 5, no. 2, pp. 143–153, 2021, doi: 10.1108/ijcs-01-2021-0001.
- [7] M. A. Ledhem, “Data mining techniques for predicting the financial performance of Islamic banking in Indonesia,” *J. Model. Manag.*, 2021, doi: 10.1108/JM2-10-2020-0286.
- [8] F. Rahman and M. I. Firdaus, “Penerapan Data Mining Metode Naïve Bayes Untuk Prediksi Hasil Belajar Siswa Sekolah Menengah Pertama (Smp),” *Al Ulum Sains dan Teknol.*, vol. 1, no. 2, pp. 76–78, 2016.
- [9] F. E. Prabowo and A. Kodar, “Analisis Prediksi Masa Studi Mahasiswa Menggunakan Algoritma Naïve Bayes,” *J. Ilmu Tek. dan Komput.*, vol. 3, no. 2, pp. 147–151, 2019, doi: 10.22441/jitkom.2020.v3.i2.008.
- [10] M. Guntur, J. Santony, and Y. Yuhandri, “Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 1, pp. 354–360, 2018, doi: 10.29207/resti.v2i1.276.
- [11] P. Meilina, “Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi,” *J. Teknol. Univ. Muhammadiyah Jakarta*, vol. 7, no. 1, pp. 11–20, 2015, [Online]. Available: jurnal.ftumj.ac.id/index.php/jurtek.
- [12] R. A. Anggraini, G. Widagdo, A. S. Budi, and M. Qomaruddin, “Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, pp. 47–51, 2019, doi: 10.26418/justin.v7i1.30211.
- [13] R. M. S. Tumangger, N. Hidayat, and Marji, “Komparasi Metode Data Mining Support Vector Machine dengan Naive Bayes untuk Klasifikasi Status Kualitas Air,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9614–9619, 2019.
- [14] B. P. Amiruddin, E. A. Kore, D. A. Ulhaq, and A. Widhatama, “Perbandingan Performa Algoritma Klasifikasi pada Data Intensitas Penggunaan Listrik Rumah Tangga,” 2020.
- [15] M. Imron, “Penerapan Data Mining Algoritma Naives Bayes Dan PART Untuk Mengetahui Minat Baca Mahasiswa Di Perpustakaan STMIK Amikom Purwokerto,” *J. Telemat.*, vol. 10, no. 2, pp. 121–135, 2017.
- [16] T. Rosandy, “Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree (C4.5) untuk Menganalisa Kelancaran Pembelajaran (Study Kasus : KSPPS / BMT AL-FADHILA),” *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 52–62, 2016.
- [17] K. S. Raju, M. R. Murty, M. V. Rao, and S. C. Satapathy, “Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction,” *Int. J. Pure Appl. Math.*, vol. 118, no. 20, pp. 321–334, 2018, [Online]. Available: https://www.researchgate.net/publication/329414359_Support_Vector_Machine_with_K-fold_Cross_Validation_Model_for_Software_Fault_Prediction.