KLASIFIKASI KASUS CACAR MONYET DI INDONESIA TAHUN 2022-2024 MENGGUNAKAN ALGORITMA NAIVE BAYES

Ahmad Farouk Adel [1], Asrul Sani [2]
[1] Sistem Informasi, Universitas Nasional
[2] Teknologi Informasi, Universitas Nasional
Jl. Sawo Manila No.61 Jakarta, Indonesia
ahmadfaroukadel2021@student.unas.ac.id

ABSTRAK

Penyakit cacar monyet merupakan salah satu penyakit menular yang kasusnya semakin meningkat di berbagai daerah. Untuk membantu dalam proses klasifikasi tingkat keparahan kasus, penelitian ini menerapkan algoritma Naïve Bayes dalam menganalisis data kasus cacar monyet di Indonesia tahun 2022–2024. Metode inidipilih karena kemampuannya dalam menangani data dengan atribut yang saling independen serta keefektifannya dalam klasifikasi. Penelitian ini menggunakan dataset yang telah dikategorikan ke dalam tiga kelas: rendah, sedang, dan tinggi. Data diolah menggunakan RapidMiner, dengan pembagian rasio data latih dan data uji yang bervariasi (50:50, 60:40, dan 70:30). Hasil evaluasi menunjukkan bahwa model dengan rasio 50:50 memberikan akurasi terbaik dibandingkan dengan rasio lainnya. Selain itu, metrik evaluasi seperti precision, recall, dan f1-score menunjukkan performa yang baik dalam mengklasifikasikan setiap kategori kasus. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan kasus cacar monyet. Temuan ini diharapkan dapat membantu instansi kesehatan dalam menentukan prioritas penanganan dan pencegahan terhadap kasus cacar monyet di Indonesia.

Kata Kunci: *Cacar Monyet, Klasifikasi, Naïve Bayes, RapidMiner, Data Mining.*

1. PENDAHULUAN

Cacar monyet (monkeypox) adalah penyakit infeksi yang disebabkan oleh virus Orthopoxvirus, yang juga mencakup virus cacar dan cacar sapi. Penyakit ini dapat menular melalui kontak langsung dengan individu terinfeksi, paparan droplet pernapasan, atau benda yang terkontaminasi, serta melalui transmisi zoonosis dari hewan ke manusia. Dalam beberapa tahun terakhir, jumlah kasus cacar monyet meningkat secara global, termasuk di Indonesia [1].

Kasus pertama cacar monyet di Indonesia dilaporkan pada tahun 2022 berdasarkan data Kementerian Kesehatan. Data ini menunjukkan bahwa provinsi dengan kepadatan penduduk tinggi dan mobilitas yang intens memiliki jumlah kasus lebih besar dibandingkan daerah lain. Hal ini mengindikasikan bahwa faktor demografi dan sosial ekonomi turut berkontribusi dalam penyebaran virus [2].

Dalam penelitian ini, algoritma Naive Bayes digunakan untuk menganalisis pola penyebaran cacar monyet berdasarkan data jumlah kasus di setiap provinsi. Algoritma ini dipilih karena kemampuannya dalam melakukan klasifikasi probabilistik berdasarkan teorema Bayes, sehingga dapat mengelompokkan provinsi ke dalam kategori risiko rendah, sedang, atau tinggi. Keunggulan utama dari Naive Bayes adalah kecepatan dan akurasinya dalam menangani dataset kategori seperti nama provinsi dan jumlah kasus [3].

Penelitian ini menggunakan RapidMiner sebagai perangkat lunak untuk mengolah dan menganalisis data. Dataset yang berisi informasi provinsi dan jumlah kasus diimpor ke dalam RapidMiner, lalu diproses menggunakan algoritma Naive Bayes untuk mengklasifikasikan provinsi berdasarkan tingkat risiko. Hasil analisis ini kemudian divisualisasikan dalam bentuk tabel atau grafik untuk membantu interpretasi data dan mendukung pengambilan keputusan dalam kebijakan kesehatan [4].

Selain melakukan klasifikasi, penelitian ini juga menggunakan data historis dari tahun 2022 hingga 2024 untuk membuat prediksi mengenai potensi peningkatan atau penurunan jumlah kasus di masa depan. Hasil prediksi ini dapat membantu pemerintah dan lembaga kesehatan dalam menyusun strategi mitigasi, seperti alokasi sumber daya medis, peningkatan kampanye kesadaran masyarakat, serta penguatan sistem pengawasan epidemiologi di wilayah dengan risiko tinggi [5]

Dengan pendekatan berbasis data menggunakan Naive Bayes dan RapidMiner, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengendalian cacar monyet di Indonesia. Hasil analisis yang diperoleh dapat menjadi dasar dalam menentukan prioritas intervensi kesehatan di provinsi yang paling terdampak, serta memberikan wawasan yang lebih mendalam mengenai pola penyebaran penyakit ini. Implementasi metode ini juga dapat diterapkan pada penelitian epidemiologi lainnya untuk memahami dan mengatasi penyakit menular di masa depan [1], [2].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

"A Review on Monkeypox" membahas monkeypox sebagai penyakit zoonosis yang kembali muncul dengan meningkatnya kasus di luar wilayah

endemik. Penelitian ini bertujuan mengkaji karakteristik, penyebaran, virologi, epidemiologi, diagnosis, manajemen, dan pencegahannya melalui kajian literatur. Hasilnya menunjukkan monkeypox disebabkan oleh Orthopoxvirus dengan dua klade utama yang berbeda tingkat virulensinya. Penyebarannya terjadi melalui kontak langsung dengan hewan atau manusia terinfeksi. Kelebihan penelitian ini adalah tinjauan komprehensifnya, sedangkan kekurangannya adalah keterbatasan data mengenai efektivitas antiviral dan pengobatan spesifik.[1]

Dalam "Public Sentiment on the Global Outbreak of Monkeypox" meneliti sentimen masyarakat terkait wabah monkeypox melalui analisis 352.182 tweet menggunakan BERT dan BERTopic. Penelitian ini menemukan tiga tema utama: kekhawatiran terhadap keselamatan pribadi, stigma terhadap komunitas minoritas, dan ketidakpercayaan pada institusi publik. Kelebihan penelitian ini adalah pemanfaatan analisis data besar secara real-time, sedangkan kekurangannya adalah keterbatasan pada tweet berbahasa Inggris serta kemungkinan perubahan persepsi seiring perkembangan wabah.[6]

Dalam "Monkeypox Global Research: A Comprehensive Analysis from Emergence to Present (1961-2023)" menganalisis tren penelitian monkeypox menggunakan bibliometrik terhadap 2.655 artikel dari Scopus. Hasilnya menunjukkan peningkatan publikasi setelah wabah terbaru, dengan Amerika Serikat dan India sebagai negara paling produktif. Tren penelitian mencakup topik seperti transmisi, pandemi, dan deep learning. Kelebihan penelitian ini adalah analisis komprehensifnya, sementara kekurangannya adalah bias dalam pemilihan artikel serta kurangnya eksplorasi aspek klinis dan epidemiologis secara mendalam.[7]

2.2. Penggunaan Data Mining dalam Epidemiologi

Data mining dalam epidemiologi adalah proses analisis data besar untuk mengidentifikasi pola dan faktor risiko yang memengaruhi kesehatan masyarakat. Dengan teknik seperti pembelajaran mesin, peneliti dapat menemukan hubungan yang tidak terlihat dalam analisis tradisional.

Metode ini digunakan untuk menganalisis rekam medis, mengidentifikasi faktor penyakit, dan memprediksi risiko berdasarkan karakteristik individu. Hasilnya mendukung penelitian epidemiologi dan pengambilan keputusan berbasis bukti dalam kebijakan kesehatan. [8]

2.3. Naïve Bayes

Naïve Bayes adalah metode klasifikasi yang didasarkan pada teorema *Bayes* dengan asumsi independensi antara fitur-fitur dalam dataset. Metode ini sering digunakan karena kemampuannya yang sederhana namun efektif dalam menganalisis data berlabel[9]

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} (1)$$

2.4. Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier (NBC) adalah metode klasifikasi berbasis probabilitas yang digunakan untuk memprediksi kelas suatu data berdasarkan atribut yang dimilikinya. Metode ini didasarkan pada Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen (asumsi "naïve"), meskipun dalam praktiknya tidak selalu demikian.

Proses kerja NBC melibatkan perhitungan probabilitas prior dari setiap kelas, diikuti dengan perhitungan probabilitas bersyarat untuk setiap atribut terhadap kelas tersebut. Kemudian, probabilitas-probabilitas ini dikombinasikan untuk menentukan kelas dengan probabilitas tertinggi.[10]

NBC telah banyak diterapkan dalam berbagai bidang, termasuk kesehatan, terutama dalam klasifikasi penyakit. Beberapa penelitian menunjukkan bahwa metode ini memiliki akurasi yang tinggi, seperti dalam klasifikasi Infeksi Saluran Pernapasan Akut (ISPA) yang mencapai akurasi 93,33% (Saputra dkk., 2023). Dengan kemampuannya dalam menangani data berukuran besar serta kecepatan pemrosesan yang tinggi, NBC menjadi salah satu metode yang efektif untuk analisis data epidemiologi dan pengambilan keputusan berbasis data.[11]

2.5. Phyton Untuk Perhitungan Klasifikasi Naïve Bayes

Python merupakan bahasa pemrograman yang banyak digunakan dalam analisis data dan machine learning, termasuk klasifikasi Naïve Bayes. Google Colab sering dimanfaatkan sebagai platform untuk menjalankan kode secara cloud tanpa perlu instalasi lokal. Beberapa library utama yang digunakan dalam penerapan Naïve Bayes meliputi Pandas untuk pengelolaan data, NumPy untuk operasi matematis, dan Scikit-learn yang menyediakan implementasi Naïve Bayes. Dalam analisis teks, TF-IDF digunakan untuk mengubah data teks menjadi bentuk numerik berdasarkan frekuensi kata.

Setelah data diproses, dataset dibagi menjadi data latih dan data uji, di mana model Naïve Bayes dibangun menggunakan data latih dan diuji dengan data uji untuk mengukur performanya. Evaluasi model dilakukan dengan metrik seperti akurasi, precision, recall, dan f-measure untuk menilai efektivitas klasifikasi. Dengan berbagai pustaka dan alat yang tersedia, Python menjadi solusi efektif dalam penerapan Naïve Bayes untuk analisis data dan pengembangan model machine learning[5].

2.6. Evaluasi Model Klasifikasi

Pengukuran terhadap kinerja suatu proses pengklasifikasian merupakan hal yang penting. Kinerja proses klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Pengukuran kinerja proses tersebut dapat menggunakan confusion matrix. Confusion matrix

adalah sebuah matriks yang memuat data klasifikasi yang dilakukan oleh sistem klasifikasi baik secara aktual maupun prediktif.[12] Confusion matrix berbentuk tabel matriks dengan kinerja model klasifikasi pada serangkaian data testing yang nilainya diketahui yang dijelaskan dalam tabel berikut:

Tabel 1.Confusian Matrix

	Actual Data		
Predicted Data		*positive*	*Negative*
	Positive	*(TP)*	*(FP)*
	Negative	*(FN)*	*(TN)*

Keterangan:

- True positive (TP), data yang kelas aktualnya adalah kelas (+) terdeteksi benar kelas (+).
- False negative (FN), data yang kelas aktualnya adalah kelas (+) namun terdeteksi kelas (-).
- False positive (FP), data yang kelas aktualnya adalah kelas (-) namun terdeteksi kelas (+).
- True negative (TN), data yang kelas aktualnya adalah kelas (-) terdeteksi benar kelas (-).

3. METODE PENELITIAN

Dalam desain penelitian ini, yang menggunakan pendekatan kuantitatif. Pendekatan kuantitatif dipilih karena data penelitian ini berbentuk numerik, serta untuk memberikan hasil analisis yang objektif dan terukur.

3.1. Fokus Penelitian

Penelitian ini berfokus pada analisis jumlah dan distribusi kasus cacar monyet di Indonesia selama tiga tahun, yaitu dari 2022 hingga 2024. Penelitian ini juga bertujuan untuk mengidentifikasi pola perubahan jumlah kasus yang terjadi setiap tahun. Berdasarkan data yang diperoleh, peneliti akan melakukan klasifikasi tingkat risiko kasus cacar monyet di setiap tahunnya. Penelitian ini diharapkan dapat memberikan wawasan yang berguna mengenai tren peningkatan atau penurunan kasus, serta perubahan pola penyebaran penyakit di berbagai wilayah di Indonesia.

3.2. Sumber data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari:

- Kementerian Kesehatan Republik Indonesia: Memberikan data resmi terkait jumlah dan sebaran kasus cacar monyet di Indonesia setiap tahun.

Data tersebut mencakup jumlah kasus yang tercatat pada tahun 2022, 2023, dan 2024, serta faktor risiko yang dapat mempengaruhi klasifikasi dan tingkat risiko di masing-masing tahun.

3.3. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan metode desk research untuk memperoleh data dari berbagai sumber yang tersedia secara publik. Teknik pengumpulan data mencakup:

a. Pengumpulan Data Sekunder: Data diperoleh dengan mengakses laporan epidemiologi cacar monyet dari sumber-sumber resmi yang dipublikasikan oleh instansi kesehatan nasional dan internasional.
b. Kompilasi Data Tahunan: Mengumpulkan dan menyusun data dari masing-masing tahun (2022, 2023, dan 2024) sehingga memungkinkan peneliti untuk menganalisis tren jumlah kasus dan perubahan yang terjadi dari tahun ke tahun.
c. Transformasi Data: Data yang telah dikumpulkan disesuaikan formatnya agar kompatibel dengan kebutuhan analisis di RapidMiner, termasuk mengonversi data kategorikal ke dalam bentuk numerik sehingga dapat diolah menggunakan algoritma Naive Bayes.

3.4. Tahap Penelitian

Beberapa tahapan menjadi kerangka penelitian dalam penelitian ini. Gambar 1 memberikan rincian lengkap tahapan penelitian.

Gambar 1. Tahapan penelitian

Berikut ringkasan dari isi Gambar 1:

a. Identifikasi Masalah dan Tujuan
   Masalah: Tren kasus cacar monyet di Indonesia dari 2022–2024 di tiap provinsi.
   Tujuan:
   - Menganalisis tren tahunan per provinsi.
   - Menggunakan Naive Bayes untuk klasifikasi tren kasus.
   - Memberikan wawasan bagi pengambilan keputusan kesehatan.
b. Pengumpulan Data
   Sumber: Data dari Kemenkes RI atau laporan epidemiologi.
   Ekstraksi: Mengunduh dalam format CSV/Excel atau scraping jika diperlukan.
   Struktur: Provinsi, Tahun, Jumlah Kasus, dan variabel epidemiologis opsional.
c. Preprocessing Data
   - Menghapus data kosong/tidak relevan.
   - Transformasi data ke format numerik.
   - Normalisasi jumlah kasus untuk keseimbangan skala data.
d. Implementasi di RapidMiner
   - Impor data dalam CSV/Excel.
   - Gunakan Naive Bayes untuk klasifikasi tren kasus.
   - Tahapan:
   - Data Splitting: Pisah data training & testing.

- Training: Model dilatih dengan data 2022–2023.
- Prediksi: Tren 2024 dianalisis.
- Visualisasi: Diagram/grafik untuk interpretasi hasil.

e. Analisis dan Interpretasi Hasil
- Mengelompokkan provinsi: Peningkatan, Penurunan, Stabil.
- Analisis faktor risiko (opsional).
- Identifikasi provinsi dengan risiko tinggi.

f. Kesimpulan dan Rekomendasi
- Ringkasan pola kasus 2022–2024 dan efektivitas Naive Bayes.
- Rekomendasi tindakan preventif di provinsi berisiko.
- Saran penelitian lanjutan dengan algoritma atau variabel tambahan.

3.5. Desain Penelitian

Desain penelitian ini menggunakan pendekatan kuantitatif dengan algoritma Naive Bayes untuk klasifikasi tingkat risiko cacar monyet. Penelitian ini bertujuan untuk mengidentifikasi pola penyebaran dan tingkat risiko penyakit cacar monyet berdasarkan data jumlah kasus dan faktor epidemiologis yang diamati dalam periode 2022–2024. Naive Bayes dipilih karena efisiensinya dalam analisis klasifikasi, terutama pada data epidemiologi.

4. HASIL DAN PEMBAHASAN

4.1. Deskriptif Dataset

Pengumpulan data didapatkan dari observasi media internet, yang menyediakan data cacar monyet, yang berjumlah 279 yang dibagi menjadi kategori, yaitu kata yang mengandung tahun, provinsi, kota, dan jumlah kasus seperti table berikut

Tabel 2. Isi Dataset

Tahun	Provinsi	Kota	Jumlah Kasus	Kategori
2022	Provinsi Banda Aceh	Banda Aceh	11	Sedang
2022	Provinsi Langsa	Langsa	11	Sedang
2022	Provinsi Lhokseumawe	Lhokseumawe	14	Tinggi
2022	Provinsi Sabang	Sabang	19	Tinggi
2022	Provinsi Subulussalam	Subulussalam	11	Sedang
2023	Provinsi Banda Aceh	Banda Aceh	11	Sedang
2023	Provinsi Langsa	Langsa	16	Tinggi
2023	Provinsi Lhokseumawe	Lhokseumawe	19	Tinggi
2023	Provinsi Sabang	Sabang	16	Tinggi
2023	Provinsi Subulussalam	Subulussalam	15	Tinggi
2024	Provinsi Banda Aceh	Banda Aceh	13	Tinggi
2024	Provinsi Langsa	Langsa	16	Tinggi
2024	Provinsi Lhokseumawe	Lhokseumawe	16	Tinggi
2024	Provinsi Sabang	Sabang	19	Tinggi
2024	Provinsi Subulussalam	Subulussalam	19	Tinggi

Tahap penerapan model Naïve Bayes dilakukan melalui beberapa pendekatan untuk memastikan keakuratan hasil serta kemudahan implementasi dalam dunia nyata. Proses ini mencakup penerapan menggunakan Python, perhitungan manual untuk validasi, serta implementasi di RapidMiner.

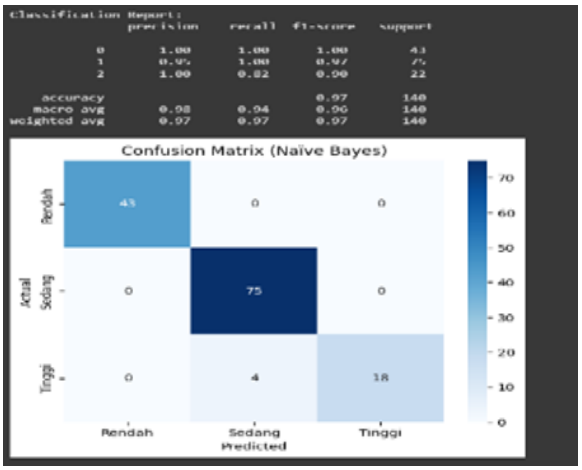
4.2. Implementasi Dengan Python

Model Naïve Bayes dikembangkan menggunakan pustaka sklearn. Langkah-langkah utama dalam implementasi ini meliputi:

a. Memuat dan mempersiapkan dataset, termasuk preprocessing data seperti normalisasi dan menangani nilai yang hilang.
b. Membagi dataset menjadi data latih dan data uji untuk memastikan model diuji dengan data baru.
c. Melatih model Naïve Bayes menggunakan GaussianNB untuk menghasilkan probabilitas klasifikasi.
d. Evaluasi model dengan confusion_matrix dan classification_report untuk melihat akurasi, precision, recall, dan F1-score.

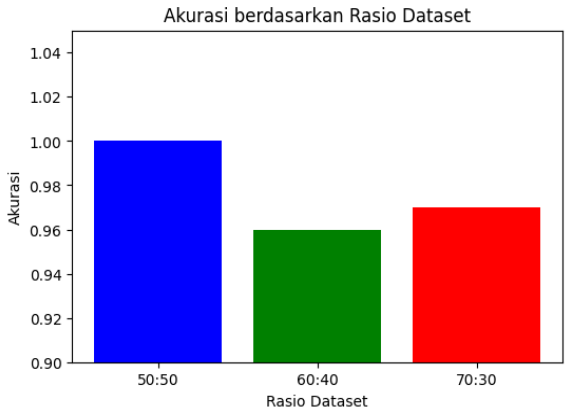
Kode utama yang digunakan:

```
fromsklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
import pandas as pd

# Memuat dataset dan preprocessing
data = pd.read_csv('dataset.csv')
X = data.drop(columns=['kelas'])
y = data['kelas']

# Membagi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Melatih model Naïve Bayes
model = GaussianNB()
model.fit(X_train, y_train)

# Evaluasi model
y_pred = model.predict(X_test)
akurasi = accuracy_score(y_test, y_pred)
print(f'Akurasi Model: {akurasi * 100:.2f}%')
```



Gambar 2. Hasil Confusion Matrix Python

Hasil evaluasi menunjukkan bahwa model memiliki akurasi keseluruhan 97%, dengan precision, recall, dan F1-score yang menunjukkan performa tinggi dalam klasifikasi.



Gambar 3 Visualisasi Pembagian Data Test

Gambar ini menunjukkan pengaruh rasio pembagian dataset terhadap akurasi model Naïve Bayes dalam mengklasifikasikan kasus cacar monyet. Sumbu X menunjukkan tiga skenario rasio pembagian dataset antara data latih dan data uji:
50:50 (50% untuk pelatihan, 50% untuk pengujian)
60:40 (60% pelatihan, 40% pengujian)
70:30 (70% pelatihan, 30% pengujian)

Sumbu Y menunjukkan nilai akurasi yang diperoleh pada masing-masing rasio pembagian dataset.
Analisis:
Rasio 50:50 (warna biru) menghasilkan akurasi tertinggi (mendekati 100%). Ini menunjukkan bahwa dengan pembagian data yang lebih seimbang antara pelatihan dan pengujian, model dapat belajar dan menggeneralisasi dengan lebih baik.

Rasio 60:40 (warna hijau) menunjukkan sedikit penurunan akurasi dibandingkan 50:50, yang mungkin disebabkan oleh berkurangnya jumlah data uji, sehingga model tidak diuji dengan cukup banyak variasi data.

Rasio 70:30 (warna merah) memiliki akurasi lebih tinggi dibandingkan 60:40 tetapi masih lebih rendah dari 50:50. Hal ini menunjukkan bahwa meskipun lebih banyak data digunakan untuk pelatihan, kinerja model tidak selalu meningkat secara signifikan.

Kesimpulannya, rasio pembagian dataset berpengaruh terhadap performa model, di mana rasio 50:50 memberikan hasil paling optimal dalam penelitian ini. Namun, rasio optimal bisa bervariasi tergantung pada jumlah data yang tersedia dan kompleksitas masalah yang dianalisis.

4.3. Perhitungan Manual

Perhitungan manual dilakukan untuk membandingkan hasil model dengan metode perhitungan probabilitas Naïve Bayes. Berdasarkan confusion matrix:

Tabel 3 Hasil Confusion Matrix

Actual / Predicted	Rendah (0)	Sedang (1)	Tinggi (2)	Total
Rendah (0)	43 (TP)	0 (FN)	0 (FN)	43
Sedang (1)	0 (FN)	75 (TP)	0 (FN)	75
Tinggi (2)	0 (FN)	4 (FP)	18 (TP)	22
Total	43	79	18	140

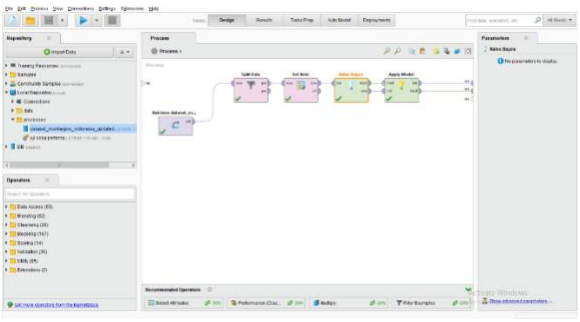
Hasil perhitungan manual menunjukkan:

Akurasi = (43 + 75 + 18) / 140 = 97.14%
Precision: Rendah = 1.00, Sedang = 0.95, Tinggi = 1.00
Recall: Rendah = 1.00, Sedang = 1.00, Tinggi = 0.82
F1-score: Rendah = 1.00, Sedang = 0.97, Tinggi = 0.90

Hasil ini menunjukkan bahwa model bekerja dengan baik dalam klasifikasi, meskipun kategori tinggi masih memiliki kelemahan dalam recall

4.4. Implementasi Dengan RapidMiner

RapidMiner digunakan sebagai metode validasi tambahan untuk membandingkan hasil model dengan metode lainnya. Langkah-langkah yang dilakukan meliputi:



Gambar 4. Tampilan Proses RapidMiner

a. Memasukkan dataset dan konfigurasi atribut target.
b. Melatih model dengan operator Naïve Bayes dan melakukan validasi silang.
c. Menganalisis hasil klasifikasi menggunakan confusion matrix serta confidence visualization untuk melihat ketepatan prediksi.

Confidence visualization di RapidMiner digunakan untuk mengukur seberapa yakin model dalam menentukan klasifikasi setiap instance. Grafik confidence visualization menampilkan probabilitas prediksi untuk setiap kelas, sehingga dapat dilihat apakah model memiliki keyakinan tinggi dalam prediksinya atau masih terdapat ketidakpastian dalam beberapa kategori. Hasil menunjukkan bahwa sebagian besar instance memiliki confidence score di atas 90%, yang berarti model sangat yakin dalam memberikan klasifikasinya. Namun, untuk kategori tinggi, terdapat beberapa instance dengan confidence score yang lebih rendah, yang mengindikasikan bahwa model masih memiliki sedikit ketidakpastian dalam membedakan kategori tinggi dan sedang.

Row No.	Kategori	prediction(Kategori)	confidence(Rendah)	confidence(Sedang)	confidence(Tinggi)	Tahun	Provinsi	Kota	Jumlah Kasus
1	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Langsa	Langsa	11
2	Sedang	Sedang	0.000	0.992	0.008	2022	Provinsi Lhokseumawe	Lhokseumawe	14
3	Tinggi	Tinggi	0.000	0.000	1.000	2022	Provinsi Sabang	Sabang	19
4	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Subulussalam	Subulussalam	11
5	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Binjai	Binjai	15
6	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Tebing Tinggi	Tebing Tinggi	13
7	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Payakumbuh	Payakumbuh	16
8	Rendah	Rendah	0.997	0.003	0.000	2022	Provinsi Sawahlunto	Sawahlunto	7
9	Tinggi	Tinggi	0.000	0.000	1.000	2022	Provinsi Bukittinggi	Bukittinggi	20
10	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Dumai	Dumai	15
11	Rendah	Rendah	1.000	0.000	0.000	2022	Provinsi Jambi	Jambi	9
12	Rendah	Rendah	1.000	0.000	0.000	2022	Provinsi Palembang	Palembang	5
13	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Pagar Alam	Pagar Alam	17
14	Sedang	Sedang	0.000	1.000	0.000	2022	Provinsi Prabumulih	Prabumulih	18
15	Rendah	Rendah	1.000	0.000	0.000	2022	Provinsi Bengkulu	Bengkulu	7
16	Rendah	Rendah	1.000	0.000	0.000	2022	Provinsi Bandar Lampung	Bandar Lampung	8
17	Rendah	Rendah	1.000	0.000	0.000	2022	Provinsi Metro	Metro	6
18	Sedang	Sedang	0.009	0.991	0.000	2022	Provinsi Jakarta Selatan	Jakarta Selatan	13

Gambar 5. Hasil confidance dari RapidMiner

Kolom-Kolom yang Ditampilkan
a. Row No. → Nomor baris dalam dataset.
b. Kategori → Kategori asli dari tingkat penyebaran (Rendah, Sedang, atau Tinggi).
c. prediction(Kategori) → Hasil prediksi model Naïve Bayes untuk kategori penyebaran.
d. confidence(Rendah), confidence(Sedang), confidence(Tinggi) → Tingkat kepercayaan model dalam menentukan apakah suatu daerah termasuk dalam kategori tertentu.
e. Tahun → Tahun data diambil (2022).
f. Provinsi → Nama provinsi di Indonesia.
g. Kota → Nama kota dalam provinsi tersebut.
h. Jumlah Kasus → Jumlah kasus cacar monyet yang tercatat di daerah tersebut.

4.4.1. Analisis Hasil Prediksi

Model berhasil memprediksi kategori dengan akurasi tinggi, terlihat dari banyaknya prediksi yang sesuai dengan kategori sebenarnya.

Contoh:

Baris 1-2: Provinsi Langsa dan Lhokseumawe dikategorikan sebagai Sedang, dan model memprediksi dengan benar kategori Sedang dengan confidence 1.000.

Baris 3: Provinsi Sabang dikategorikan sebagai Tinggi, dan model memprediksi Tinggi dengan confidence 1.000.

Baris 12-17: Beberapa provinsi seperti Palembang, Pagar Alam, Prabumulih, Bengkulu, dan Bandar Lampung dikategorikan Rendah, dan model memprediksi Rendah dengan confidence 1.000.

4.4.2. Confidence Score

Nilai confidence menunjukkan tingkat keyakinan model terhadap keputusan yang dibuatnya. Misalnya, di baris ke-8 (Sawahlunto), model memberi confidence 0.997 untuk Rendah, yang sangat mendekati 1, menunjukkan kepercayaan tinggi dalam prediksinya. Di baris ke-18 (Jakarta Selatan), confidence 0.991 untuk Sedang, yang juga menunjukkan kepercayaan tinggi

Hasil evaluasi dari RapidMiner menunjukkan bahwa model mencapai akurasi yang sama dengan metode Python dan perhitungan manual, membuktikan bahwa model bekerja secara konsisten dalam berbagai platform.

5. KESIMPULAN DAN SARAN

Algoritma Naïve Bayes berhasil mengklasifikasikan tingkat penyebaran cacar monyet di Indonesia tahun 2022–2024 dengan akurasi 97%. Model memiliki presisi tinggi untuk kategori rendah (100%) dan sedang (95%), sedangkan recall kategori tinggi mencapai 82%. Hasil evaluasi menggunakan confusion matrix menunjukkan konsistensi antara implementasi di Python dan RapidMiner, dengan kesesuaian hasil di berbagai metode validasi.

Sebagian besar kasus masuk dalam kategori sedang, diikuti oleh rendah, sementara kategori tinggi memiliki jumlah lebih sedikit dan tingkat kesalahan prediksi lebih tinggi. Keunggulan model ini terletak pada kemampuannya menangani dataset kompleks serta memberikan hasil prediksi yang cepat dan akurat. Confidence visualization di RapidMiner menunjukkan tingkat keyakinan tinggi pada sebagian besar prediksi, meskipun masih terdapat ketidakpastian pada kategori tinggi. Model ini terbukti efektif dalam analisis epidemiologi dan berpotensi dikembangkan lebih lanjut dalam sistem berbasis web atau dashboard interaktif. Dengan validasi yang konsisten, model ini dapat membantu tenaga medis dan otoritas kesehatan dalam pemantauan serta pengendalian penyebaran cacar monyet secara lebih efektif.

Model ini dapat dikembangkan lebih lanjut dengan teknik balancing data atau kombinasi dengan algoritma lain untuk meningkatkan akurasi klasifikasi. Penambahan variabel seperti faktor geografis dan pola mobilitas dapat memperkaya analisis prediksi. Selain itu, implementasi model dalam sistem berbasis web akan mempermudah tenaga medis dalam mengakses

hasil klasifikasi secara real-time. Validasi dengan dataset yang lebih besar dan kerja sama dengan institusi kesehatan juga dapat dilakukan untuk meningkatkan keakuratan serta penerapan model dalam dunia nyata.akurat.

DAFTAR PUSTAKA

[1] V. Panchal, V. Gajera, T. Desai, D. Parekh, A. Professor, and A. Professor, "A Review on Monkeypox," 2023. [Online]. Available: www.ijfmr.com

[2] M. S. Syarah, M. Wati, N. Puspitasari, and S. Artikel, "Klasifikasi Penderita ISPA Menggunakan Metode Naive Bayes Classifier INFORMASI ARTIKEL ABSTRACT," *INNOVATION IN RESEARCH OF INFORMATICS*, vol. 4, no. 1, pp. 8–15, 2022.

[3] Y. Muhyidin and M. R. Muttaqin, "Penerapan Algoritma Naïve Bayes Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapid Miner Application Of The Naïve Bayes Algorithm For Prediction Of Lung Diseases Using Rapidminer," vol. 2024, no. 1, pp. 145–152, 2024, doi: 10.51132/teknologika.v14i1.

[4] T. N. padilah Padilah and R. I. Adam, "ANALISIS REGRESI LINIER BERGANDA DALAM ESTIMASI PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG," 2019.

[5] A. P. Sukma Wahyu *et al.*, "PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFICATION UNTUK KLASIFIKASI SENTIMENT TWEET TERHADAP PLATFORM STREAMING ILEGAL," *Jurnal informasi dan Komputer*, vol. 11, no. 2, p. 2023, 2023.

[6] Q. X. Ng, C. E. Yau, Y. L. Lim, L. K. T. Wong, and T. M. Liew, "Public sentiment on the global outbreak of monkeypox: an unsupervised machine learning analysis of 352,182 twitter posts," *Public Health*, vol. 213, pp. 1–4, Dec. 2022, doi: 10.1016/j.puhe.2022.09.008.

[7] N. Kameli *et al.*, "Monkeypox Global Research: A Comprehensive Analysis from Emergence to Present (1961-2023) for innovative prevention and control approaches," Jan. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.jiph.2024.102593.

[8] I. N. B. gunawan, N. Putri Lestari, and R. Dwi Kurniawan, "TINJAUAN PUSTAKA SISTEMATIS: DATA MINING DALAM BIDANG KESEHATAN," 2022. [Online]. Available: https://jetbis.al-makkipublisher.com/index.php/al/index

[9] A. S. R. Fauziah, "Penggunaan Rapidminer Untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes," Jun. 2024.

[10] W. Anugrah, E. Haerani, and L. Oktavia, "Klasifikasi Penyakit Cacar Monyet Menggunakan Metode Support Vector Machine," 2024, doi: 10.47065/josyc.v5i3.5149.

[11] T. O. Saputra, D. Alamsyah, and K. Kunci, "2 ND MDP STUDENT CONFERENCE (MSC) 2023 Universitas Multi Data Palembang | KLASIFIKASI PENYAKIT CACAR MONYET MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK," 2023.

[12] Deny Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," vol. 21, no. 1, 2019, doi: 10.31294/p.v20i2.