

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Cacar monyet, atau dikenal sebagai monkeypox, adalah penyakit infeksi yang disebabkan oleh virus DNA ganda dari genus Orthopoxvirus, yang juga mencakup virus cacar dan cacar sapi. Penyakit ini menjadi perhatian global karena pola penyebarannya yang kompleks dan meningkatnya jumlah kasus dalam beberapa tahun terakhir. Virus ini dapat menular melalui berbagai jalur, seperti kontak langsung dengan individu yang terinfeksi, paparan droplet pernapasan, atau melalui benda-benda yang terkontaminasi virus, seperti pakaian atau tempat tidur. Selain itu, transmisi zoonosis, yaitu dari hewan ke manusia, juga berkontribusi pada penyebaran virus ini, terutama di wilayah yang memiliki interaksi tinggi antara manusia dan hewan liar (Panchal dkk., 2023).

Di Indonesia, laporan pertama terkait cacar monyet muncul pada tahun 2022 dalam data Kementerian Kesehatan, yang menunjukkan distribusi kasus berdasarkan nama provinsi dan jumlah kasus yang terdeteksi. Informasi ini sangat penting untuk memahami pola penyebaran virus dan mengidentifikasi wilayah yang memerlukan intervensi kesehatan lebih lanjut. Berdasarkan data tersebut, diketahui bahwa beberapa provinsi dengan kepadatan penduduk tinggi dan mobilitas yang intens memiliki jumlah kasus yang lebih besar dibandingkan provinsi lainnya. Hal ini menunjukkan bahwa faktor demografi dan sosial ekonomi turut memengaruhi distribusi penyakit ini (Syarah dkk., 2022).

Seiring dengan meningkatnya perhatian terhadap cacar monyet, diperlukan pendekatan berbasis data untuk menganalisis pola penyebarannya. Data yang dikumpulkan dari situs Kementerian Kesehatan tentang jumlah kasus di setiap provinsi dapat menjadi dasar untuk analisis lebih lanjut. Penelitian ini bertujuan untuk memanfaatkan algoritma Naive Bayes dalam mengklasifikasikan data tersebut,

sehingga dapat memberikan wawasan tentang risiko penyebaran di masa depan serta membantu pengambilan keputusan yang lebih tepat sasaran (Syarah dkk., 2022).

Penelitian ini menggunakan algoritma naïve bayes untuk Analisa data karena Algoritma Naive Bayes adalah metode klasifikasi probabilistik yang sangat cocok untuk analisis data berbasis kategori seperti nama provinsi dan jumlah kasus. Metode ini didasarkan pada teorema Bayes, yang menghitung probabilitas suatu kejadian berdasarkan informasi sebelumnya. Dalam konteks ini, Naive Bayes dapat digunakan untuk mengklasifikasikan provinsi berdasarkan jumlah kasus menjadi beberapa kategori, seperti risiko rendah, sedang, atau tinggi. Keunggulan utama algoritma ini adalah kemampuannya untuk memberikan hasil yang cepat dan akurat, meskipun dengan dataset yang sederhana (Muhyidin & Muttaqin., 2024).

Naive Bayes bekerja dengan asumsi bahwa setiap variabel dalam dataset bersifat independen satu sama lain. Dalam kasus cacar monyet, variabel yang dianalisis meliputi nama provinsi sebagai kategori, dan jumlah kasus sebagai atribut numerik. Dengan pendekatan ini, algoritma Naive Bayes dapat mengidentifikasi pola distribusi kasus di setiap provinsi, yang kemudian dapat digunakan untuk menentukan prioritas intervensi kesehatan. Misalnya, provinsi dengan jumlah kasus tinggi dapat dianggap sebagai wilayah risiko tinggi yang memerlukan perhatian lebih intensif dari pemerintah dan lembaga kesehatan masyarakat.

Selain itu, algoritma Naive Bayes juga dapat digunakan untuk membuat prediksi berdasarkan pola data historis. Dalam penelitian ini, data dari tahun 2022 hingga 2024 akan dianalisis untuk memperkirakan kemungkinan peningkatan atau penurunan jumlah kasus di masa mendatang. Hasil prediksi ini dapat membantu pihak berwenang dalam merencanakan langkah-langkah mitigasi, seperti alokasi sumber daya kesehatan, peningkatan kampanye kesadaran masyarakat, dan penguatan sistem pengawasan penyakit di provinsi dengan risiko tinggi (Syarah dkk., 2022).

Penuulisan Ini memanfaatkan Rapid Miner untuk Untuk mendukung analisis data dan penerapan algoritma Naive Bayes, RapidMiner adalah sebuah perangkat lunak open-source yang dirancang untuk data mining, analisis data, dan pengembangan model prediktif. RapidMiner memiliki antarmuka yang ramah pengguna, sehingga memudahkan peneliti untuk mengolah data yang kompleks. Dalam konteks penelitian ini, data mengenai nama provinsi dan jumlah kasus cacar monyet akan diolah menggunakan RapidMiner sebagai input utama untuk analisis (Padilah & Adam, 2019).

Proses analisis dalam RapidMiner dimulai dengan mengimpor dataset yang berisi informasi tentang provinsi dan jumlah kasus. Setiap baris dalam dataset mewakili satu provinsi, sementara kolomnya berisi atribut, seperti nama provinsi dan jumlah kasus yang terdeteksi. Setelah data diimpor, algoritma Naive Bayes diterapkan untuk mengklasifikasikan provinsi berdasarkan tingkat risiko. Proses ini dilakukan melalui design perspective, yaitu antarmuka kerja utama dalam RapidMiner, yang memungkinkan pengguna untuk merancang alur kerja analisis secara visual.

Hasil analisis kemudian ditampilkan dalam result perspective, yang menyediakan berbagai opsi visualisasi, seperti grafik atau tabel, untuk memudahkan interpretasi. Dengan memanfaatkan fitur-fitur ini, penelitian dapat menghasilkan laporan yang jelas dan informatif tentang distribusi kasus cacar monyet di seluruh Indonesia. Selain itu, hasil analisis juga dapat digunakan untuk menyusun rekomendasi kebijakan kesehatan yang lebih spesifik, seperti memperkuat edukasi tentang pencegahan cacar monyet di provinsi dengan risiko tinggi (Padilah & Adam, 2019).

Dengan Demikian Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam upaya pengendalian cacar monyet di Indonesia. Dengan menggunakan data dari situs Kementerian Kesehatan, algoritma Naive Bayes, dan perangkat lunak RapidMiner, penelitian ini menawarkan pendekatan berbasis data untuk memahami pola penyebaran penyakit. Hasil klasifikasi yang dihasilkan dapat

membantu pihak berwenang dalam menentukan wilayah prioritas untuk intervensi kesehatan, sekaligus memberikan alat prediksi untuk mengantisipasi penyebaran virus di masa depan (Panchal dkk., 2023; Syarah dkk., 2022;).

Sebagai contoh, jika data menunjukkan bahwa provinsi tertentu secara konsisten memiliki jumlah kasus tinggi, langkah-langkah seperti penguatan fasilitas kesehatan, peningkatan edukasi masyarakat, dan pengawasan yang lebih ketat dapat segera dilakukan. Sebaliknya, untuk provinsi dengan jumlah kasus rendah, fokus dapat diberikan pada pencegahan, seperti kampanye kesadaran masyarakat untuk mengurangi risiko penularan dari hewan ke manusia. Dengan demikian, hasil penelitian ini tidak hanya memberikan gambaran tentang situasi terkini, tetapi juga berkontribusi pada pengambilan keputusan strategis dalam pengendalian penyakit di masa mendatang (Padilah & Adam, 2019).

Oleh karena itu disimpulkan dari latar belakang jurnal penelitian ini bahwa cacar monyet ini adalah tantangan kesehatan global yang membutuhkan pendekatan berbasis data untuk pengendalian yang efektif. Dengan menggunakan data nama provinsi dan jumlah kasus dari situs Kementerian Kesehatan, algoritma Naive Bayes dapat memberikan wawasan yang berharga tentang distribusi risiko di Indonesia. Melalui implementasi dalam RapidMiner dan python di google colab, penelitian ini mampu menghasilkan klasifikasi yang akurat dan prediksi yang relevan, yang dapat membantu pihak berwenang dalam merencanakan langkah-langkah mitigasi yang tepat.

Penelitian ini menunjukkan bahwa penggabungan teknologi analisis data dengan algoritma yang tepat dapat memberikan dampak yang signifikan dalam memahami dan mengatasi tantangan kesehatan masyarakat. Oleh karena itu, penelitian ini diharapkan dapat menjadi landasan bagi upaya pencegahan dan pengendalian cacar monyet di Indonesia, sekaligus mendorong penggunaan teknologi serupa untuk mengatasi penyakit menular lainnya di masa depan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah penelitian ini adalah sebagai berikut:

- Bagaimana model klasifikasi yang dibangun menggunakan algoritma Naive Bayes dapat mengidentifikasi tingkat risiko kasus cacar monyet di Indonesia tahun 2025?
- Seberapa akurat model Naive Bayes dalam mengklasifikasikan data kasus cacar monyet berdasarkan faktor epidemiologi dan demografi?

## 1.3 Tujuan Penelitian

Tujuan penelitian ini adalah:

- Membangun model klasifikasi kasus cacar monyet menggunakan algoritma Naive Bayes yang diimplementasikan dalam RapidMiner.
- Mengevaluasi performa model klasifikasi Naive Bayes dalam mengidentifikasi tingkat risiko kasus cacar monyet di Indonesia pada tahun 2022-2024

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya dalam bidang data mining dan analisis prediktif menggunakan algoritma Naive Bayes untuk kasus epidemiologi.
- Bagi pemerintah dan instansi kesehatan, hasil penelitian ini dapat digunakan sebagai salah satu acuan dalam melakukan pemantauan, pencegahan, dan pengendalian kasus cacar monyet di Indonesia. Model klasifikasi yang dihasilkan dapat membantu pihak berwenang dalam mengalokasikan sumber daya kesehatan secara lebih efektif ke wilayah-wilayah yang memiliki tingkat risiko lebih tinggi.

- Dengan klasifikasi yang akurat, penelitian ini dapat menjadi dasar bagi kebijakan kesehatan yang lebih terfokus pada pengendalian penyakit zoonosis, termasuk cacar monyet, dalam konteks pencegahan epidemi yang lebih luas.

### 1.5 Batasan Penelitian

Penelitian ini memiliki beberapa batasan, di antaranya:

- Sumber data yang digunakan dalam penelitian ini berasal dari laporan institusi kesehatan dan sumber epidemiologi resmi, sehingga terdapat kemungkinan keterbatasan informasi mengenai faktor-faktor lain yang mungkin relevan dalam klasifikasi kasus cacar monyet.
- Naive Bayes memiliki asumsi bahwa setiap fitur independen satu sama lain, yang mungkin tidak sepenuhnya akurat dalam konteks kasus cacar monyet, mengingat adanya faktor-faktor yang dapat saling berkaitan.
- Google Colab memiliki keterbatasan RAM, GPU, dan penyimpanan, yang dapat mempengaruhi pemrosesan dataset besar. Sesi Google Colab juga memiliki batas waktu, sehingga proses training model dapat terputus. Selain itu, penggunaan Colab memerlukan koneksi internet yang stabil, serta bergantung pada pustaka yang tersedia seperti Scikit-Learn, Pandas, dan Matplotlib.
- RapidMiner dipilih sebagai perangkat analisis. Hasil yang didapatkan mungkin memiliki keterbatasan dibandingkan dengan hasil yang diperoleh dari perangkat analisis lain.