

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Machine Learning (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk meniru dan menggantikan peran manusia untuk memecahkan masalah[1]. Dalam konsep tradisional pemograman, akan memberikan aturan dan instruksi kepada computer untuk melakukan sesuatu tugas tertentu[2]. ML memiliki focus pada pengembangan sebuah system yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali deprogram oleh manusia. Metode tersebut, mesin tidak hanya bisa menemukan aturan untuk perilaku optimal dalam pengambilan keputusan, namun juga bisa beradaptasi dengan perubahan yang terjadi[3]. Selain itu, dalam proses ML ini menggunakan metode klasifikasi.

Klasifikasi merupakan salah satu teknik dalam *machine learning* yang bertujuan untuk menyusun data ke dalam kelompok atau kelas tertentu berdasarkan atribut atau ciri-ciri yang dimilikinya. Proses ini berfokus pada upaya untuk memprediksi kelas dari data yang belum diketahui sebelumnya dengan memanfaatkan pola yang telah dipelajari dari data historis. Dengan memilih algoritma klasifikasi yang sesuai, berbagai bidang dapat memperoleh peningkatan dalam efisiensi serta ketepatan dalam proses pengambilan keputusan [4].

Ada banyak teknik klasifikasi yang dapat digunakan dalam bidang machine learning, di antaranya adalah *Naive Bayes*. algoritma ini telah banyak digunakan oleh peneliti-peneliti sebelumnya dalam berbagai aplikasi seperti klasifikasi data medis, pengenalan pola, deteksi penyakit, dan prediksi diagnosis klinis [5].

Naive Bayes adalah algoritma klasifikasi berbasis probabilistik yang menerapkan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen atau bebas secara statistik satu sama lain. Meskipun asumsi ini sering kali tidak sepenuhnya sesuai dengan data dunia nyata, *Naive Bayes* terbukti efektif dan efisien dalam banyak kasus, terutama ketika diterapkan pada data berukuran besar. Algoritma ini sangat cepat dalam proses pelatihan dan prediksi, serta cukup andal untuk klasifikasi awal dalam sistem deteksi penyakit .

Naive Bayes telah digunakan dalam berbagai studi klasifikasi penyakit, termasuk penyakit menular. Dalam konteks penelitian ini, algoritma Naive Bayes digunakan untuk mengklasifikasikan infeksi virus cacar monyet berdasarkan gejala dan data klinis pasien. Diharapkan penerapan algoritma ini dapat memberikan hasil klasifikasi yang efektif dalam mendeteksi kasus positif maupun negatif, serta memberikan wawasan mengenai kelebihan dan keterbatasannya dalam klasifikasi penyakit.

Dataset infeksi virus cacar monyet yang digunakan dalam penelitian ini diperoleh dari *Monkeypox Coursework Dataset*, yang terdiri dari 25.000 data pasien dengan 18 atribut, serta dua kelas output, yaitu Positive (terinfeksi cacar monyet) dan Negative (tidak terinfeksi cacar monyet). Virus cacar monyet merupakan penyakit zoonosis langka yang dapat menular dari hewan ke manusia,

dan kemudian antar manusia. Gejala umumnya mencakup demam, pembengkakan kelenjar getah bening, dan ruam kulit, yang bisa sangat mirip dengan penyakit menular lainnya, sehingga membutuhkan diagnosis yang akurat dan cepat.

Infeksi virus cacar monyet dapat berkembang serius terutama pada individu dengan sistem kekebalan tubuh lemah. Gejala-gejala klinis yang ditunjukkan pasien, seperti nyeri rektal, lesi pada mulut, pembengkakan amandel, dan gejala sistemik lainnya, menjadi indikator penting dalam proses diagnosis. Sayangnya, pada fase awal infeksi, beberapa gejala tidak selalu muncul atau cukup jelas untuk dikenali, sehingga meningkatkan risiko keterlambatan diagnosis dan penyebaran penyakit.

Penelitian ini bertujuan untuk mengevaluasi secara empiris kinerja algoritma *Naive Bayes* dalam mengklasifikasikan infeksi virus cacar monyet berdasarkan gejala dan data klinis pasien. Hasil penelitian diharapkan dapat digunakan untuk mengembangkan model deteksi dini yang akurat, sehingga dapat membantu tenaga medis dalam mengambil keputusan lebih cepat dan berbasis data. Dengan sistem klasifikasi otomatis berbasis *Machine Learning*.

Sehubungan dengan latar belakang diatas, maka peneliti tertarik untuk melakukan penelitian untuk mendapatkan akurasi yang lebih baik lagi dari penelitian sebelumnya dengan berjudul “**Implementasi Algoritma *Naive Bayes* Dalam Klasifikasi Infeksi Virus Cacar Monyet**”.

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang yang telah diuraikan di atas, maka di dapat

Rumusan masalah sebagai berikut :

1. Bagaimana menerapkan algoritma *Naive bayes* untuk mengklasifikasikan infeksi virus cacar monyet?
2. Bagaimana mengevaluasi kinerja algoritma Naive Bayes dalam mengklasifikasikan infeksi virus cacar monyet berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan F1-score?

1.3 BATASAN MASALAH

Untuk menghindari permasalahan yang meluas, maka penulis membatasi pembahasan permasalahan sebagai berikut:

1. Dataset yang digunakan merupakan yang terdiri dari 25.000 data pasien dengan 18 atribut dan dua kelas output (*Positive* dan *Negative*). Data ini bersumber dari *repository public* yaitu *kaggle.com* dengan nama *Monkeypox Coursework Dataset*.
2. Metode klasifikasi serta algoritma yang digunakan yaitu *Naive bayes*.
3. Evaluasi kinerja model dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score* dengan metode validasi silang (*cross-validation*).
4. Pengelolaan data dilakukan menggunakan bahasa pemrograman Python.
5. Aplikasi yang digunakan adalah *Google Colaboratory*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Untuk menerapkan algoritma *Naive Bayes* dalam mengklasifikasikan infeksi virus cacar monyet.
2. Untuk mengevaluasi kinerja algoritma *Naive Bayes* menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4.2 Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan Penelitian ini, yaitu :

1. Memberikan informasi mengenai efektivitas algoritma *Naive Bayes* dalam klasifikasi infeksi virus cacar monyet.
2. Menjadi referensi bagi peneliti lain dalam memilih algoritma klasifikasi yang sesuai untuk penyakit menular.
3. Mendukung pengembangan sistem deteksi dini berbasis machine learning untuk membantu tenaga medis dalam proses diagnosis.

1.5 SISTEMATIKA PENULISAN

Sistematika penulisan menggambarkan secara umum tentang apa yang akan dibahas dalam setiap bab dari laporan penelitian ilmiah ini. Adapun isi pokok bahasan masing-masing bab sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, Batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini membahas teori-teori yang relevan sebagai landasan penelitian, seperti konsep dasar *machine learning*, algoritma *Naive Bayes*, serta studi terdahulu yang berkaitan dengan klasifikasi penyakit menular.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk tahapan penelitian, deskripsi *dataset*, proses *preprocessing data*, penerapan algoritma, dan metode evaluasi kinerja model..

BAB IV : HASIL DAN ANALISIS

Pada Bab ini menyajikan hasil implementasi algoritma *Naive Bayes* pada dataset infeksi virus cacar monyet. Selanjutnya dilakukan analisis kinerja model berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

BAB V : KESIMPULAN

Pada bab ini berisikan Kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.