

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan dunia kesehatan yang semakin maju serta kemudahan akses layanan kesehatan di era modern saat ini dirahapkan dapat menekan angka penyebaran berbagai jenis virus menular secara signifikan di seluruh dunia. Namun, dilapangan virus tetap menjadi salah satu ancaman kesehatan global dan belum sepenuhnya dapat dihentikan karena sifat yang sangat menular dengan kemampuannya untuk bermutasi secara cepat, tidak memungkinkan untuk menghentikan penyebaran sebuah virus, terutama virus yang memiliki sifat penyebaran yang masif seperti varian Virus Influenza [1].

Virus influenza sendiri merupakan masalah kesehatan yang menginfeksi saluran pernapasan yang dimana virus ini menjadi epidemic flu musiman pada setiap tahunnya yang di sebabkan oleh 2 tipe virus yaitu tipe A dan tipe B [2]. Secara global, virus influenza telah menimbulkan sekitar satu miliar kasus setiap tahun, dengan jumlah 3 hingga 5 juta kasus berat, dan telah mengakibatkan 290.000 hingga 650.000 kematian akibat komplikasi pernapasan [1]. Di Indonesia sendiri, total kasus influenza di perkirakan mencapai empat juta kasus per tahun dengan hampir 200.000 rawat inap [3]. Penyebaran virus Influenza sering kali terjadi pada masuknya musim penghujan atau musim dingin, kemudian Virus ini dapat di tularkan melalui udara oleh batuk dan bersin [4].

Sebagai langkah pencegahan, vaksinasi tahunan dijadikan solusi dalam menekan angka penyebaran virus influenza. Namun beberapa penelitian sebelumnya juga lebih banyak berfokus pada aspek medis dan epidemiologis, seperti efektivitas vaksin, tingkat perlindungan terhadap infeksi, serta dampaknya terhadap penurunan angka rawat inap dan kematian [2], [4]. Penelitian yang secara khusus mengkaji klasifikasi penerimaan vaksin influenza menggunakan pendekatan machine learning masih sangat terbatas. Selain itu, belum banyak kajian yang membandingkan beberapa algoritma klasifikasi untuk menentukan

metode yang paling optimal dalam memprediksi penerimaan vaksin berdasarkan faktor demografis, perilaku, dan persepsi individu. Informasi tersebut diharapkan menjadi dasar dalam menyusun strategi penanganan dan program edukasi yang lebih efisien, tepat sasaran, dan berdampak nyata terhadap ketahanan kesehatan masyarakat.

Untuk menjawab kompleksitas faktor yang memengaruhi keputusan seseorang dalam menerima vaksin, penelitian ini menerapkan teknologi *machine learning* sebagai pendekatan utama dalam proses analisis dan prediksi penerimaan vaksin. Dataset yang digunakan terdiri dari 26.707 baris data. Dalam penelitian ini digunakan tiga algoritma klasifikasi, yaitu Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Algoritma Naïve Bayes merupakan metode klasifikasi yang berlandaskan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain, sehingga mampu menangani dataset berukuran besar secara efisien dan menghasilkan prediksi dengan cepat [5]. Support Vector Machine (SVM) bekerja dengan mencari *hyperplane* optimal yang memaksimalkan jarak antar kelas dalam proses klasifikasi [6]. Sementara itu, Random Forest merupakan algoritma yang membangun sekumpulan pohon keputusan (*decision trees*) dan menggabungkan hasil prediksinya untuk memperoleh hasil yang lebih optimal, serta memiliki kemampuan yang baik dalam menangani data tidak seimbang [7].

Berdasarkan uraian diatas, diperlukan suatu pendekatan analisis yang mampu mengklasifikasikan penerimaan vaksin influenza secara akurat berdasarkan karakteristik data. Metode CRISP-DM (Cross Industry Standard Process for Data Mining) dipilih sebagai kerangka kerja sistematis untuk melakukan penelitian ini yang menerapkan algoritma Support Vector Machine (SVM), Naïve Bayes, dan Random Forest dalam klasifikasi penerimaan vaksin influenza. Penelitian ini dilakukan untuk mengetahui algoritma yang paling efektif dan akurat dalam klasifikasi data vaksinasi. Oleh karena itu, Hasil penelitian diharapkan menjadi referensi bagi peneliti selanjutnya dan diharapkan memberikan kontribusi dibidang kesehatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah di uraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

Bagaimana kinerja algoritma *Support Vector Machine* (SVM), *Naive Bayes*, dan *Random Forest* dalam mengklasifikasikan penerimaan vaksin influenza?

1.3 Batasan Masalah

Penelitian ini memiliki batasan agar lebih terarah, maka Batasan masalah adalah sebagai berikut:

1. Dalam penelitian ini menggunakan dataset terbatas yang bersumber Kaggle <https://www.kaggle.com/datasets/arashnic/flu-data>.
2. Penelitian ini hanya fokus pada algoritma *Support Vector Machine* (SVM), *Naive Bayes*, dan *Random Forest*.
3. Klasifikasi menggunakan target `h1n1_vaccine` dengan label Penerimaan Vaksin (1) dan Tidak Menerima Vaksin (0).

1.4 Tujuan Penelitian

Dalam penelitian ini tujuan yang ingin di capai adalah Menganalisis dan membandingkan kinerja ketiga algoritma (SVM, *Naive Bayes*, *Random Forest*) dalam menentukan algoritma terbaik untuk klasifikasi vaksin influenza.

1.5 Manfaat Penelitian

Dalam penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan pemahaman mengenai kinerja dan karakteristik algoritma *Naive Bayes*, *Support Vector Machine*, dan *Random Forest* dalam mengklasifikasikan penerima vaksin influenza.
2. Hasil penelitian diharapkan dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan dan menerapkan metode machine

learning, khususnya untuk klasifikasi penerimaan vaksin influenza.

1.6 Sistematika Penulisan

Berikut adalah sistematika penulisan yang diperlukan untuk Menyusun dan menyelesaikan penelitian ini:

BAB I PENDAHULUAN

Bab 1 berisi tentang latar belakang masalah yang menggambarkan kerangka pada penelitian yang akan dilakukan, diikuti dengan rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab 2 berisi tentang tinjauan pustaka yang terdiri dari studi literatur yang membahas penelitian-penelitian terdahulu yang bersumber dari jurnal-jurnal untuk membandingkan penelitian yang akan dilakukan, kemudian dasar teori yang membahas teori-teori yang relevan terhadap alur penelitian sehingga dapat memperkuat penelitian.

BAB III METODE PENELITIAN

Bab III membahas metodologi penelitian yang digunakan dalam penelitian ini, meliputi jenis dan pendekatan penelitian, alur penelitian berdasarkan tahapan CRISP-DM, serta penjelasan mengenai objek dan data penelitian yang digunakan. Pada bab ini juga diuraikan tahapan metode penelitian mulai dari business understanding, data understanding, data preparation, modeling, evaluation, hingga deployment dalam bentuk visualisasi hasil. Selain itu, dijelaskan pula teknik analisis data yang digunakan serta alat dan bahan penelitian yang mendukung proses analisis dan pemodelan algoritma machine learning.

BAB IV HASIL DAN PEMBAHASAN

Bab IV membahas mengenai hasil dari tahapan penelitian yang telah dilakukan, mulai dari deskripsi data, hasil pemrosesan awal (preprocessing), hingga hasil pengujian model. Pada bab ini dijabarkan perbandingan performa antara algoritma SVM, Naive Bayes, dan Random Forest dengan variasi test size

10%, 20%, dan 30%. Selain itu, dilakukan analisis mendalam mengenai evaluasi model menggunakan Confusion Matrix, kurva ROC, serta implementasi deployment model terbaik ke dalam aplikasi berbasis web menggunakan framework Streamlit.

BAB V PENUTUP

Bab V berisi kesimpulan dari seluruh rangkaian penelitian yang telah dilaksanakan. Kesimpulan ditarik berdasarkan hasil analisis perbandingan untuk menentukan algoritma terbaik dalam memprediksi penerimaan vaksin influenza. Selain itu, bab ini juga memuat saran-saran yang ditujukan bagi pengembangan penelitian selanjutnya serta pengembangan sistem aplikasi agar dapat memberikan kontribusi yang lebih luas di masa mendatang.

