Contents lists available at JESA

Journal of Engineering and Science Application

journal homepage: https://jesa.aks.or.id/index.php/jesa/index

Prediksi Penyakit Ispa Menggunakan Data Mining Naive Bayes di Klinik Pratama Rawat Inap HMS Medika

Riki Datul Fajar
Teknik Informatika, Universitas Pamulang

Article Info

Keywords:

Data Mining;
Naive Bayes;
ISPA;
Prediksi;
Sistem Berbasis Web

ABSTRACT

Infeksi Saluran Pernapasan Akut (ISPA) merupakan gangguan kesehatan yang perlu dikenali sejak dini karena gejalanya dapat muncul dalam beberapa bentuk, seperti demam, batuk, pilek, sesak napas, dan nyeri tenggorokan. Di Klinik Pratama Rawat Inap HMS Medika, identifikasi penyakit masih bergantung pada pemeriksaan dan analisis tenaga medis, sedangkan data pemeriksaan yang tersimpan belum dimanfaatkan secara maksimal untuk menemukan pola gejala. Penelitian ini bertujuan menerapkan data mining dengan algoritma Naive Bayes untuk memprediksi kemungkinan ISPA berdasarkan gejala pasien dan membangun aplikasi berbasis web sebagai alat bantu identifikasi awal. Pengumpulan informasi dilakukan melalui wawancara, observasi, dan studi literatur. Atribut yang digunakan mencakup umur, jenis kelamin, demam, batuk, pilek, sesak napas, dan nyeri tenggorokan, dengan kelas prediksi ISPA dan bukan ISPA. Data melalui tahap preprocessing berupa pemeriksaan kelengkapan, penanganan data tidak lengkap, pemilihan atribut, dan pengelompokan berdasarkan kelas. Aplikasi dikembangkan menggunakan PHP, framework CodeIgniter, dan basis data MySQL. Perancangan sistem memakai Unified Modeling Language (UML), pengujian fungsi memakai Black Box Testing, dan evaluasi klasifikasi direncanakan memakai confusion matrix serta akurasi. Hasil penelitian menunjukkan bahwa sistem mampu menerima data gejala, menghitung peluang setiap kelas, dan menampilkan hasil prediksi ISPA atau bukan ISPA. Data yang digunakan berupa data dummy dan simulasi, sedangkan nilai akurasi belum dilaporkan, sehingga kinerja klasifikasi belum dapat dinyatakan secara kuantitatif. Sistem diposisikan sebagai alat bantu dan tidak menggantikan pemeriksaan klinis atau diagnosis dokter.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Riki Datul Fajar
Teknik Informatika
Universitas Pamulang
Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
Email: rikifajar903@gmail.com

1. PENDAHULUAN

Infeksi Saluran Pernapasan Akut (ISPA) merupakan masalah kesehatan yang berkaitan dengan gangguan pada saluran pernapasan. Gejala yang dialami pasien dapat berbeda-beda, antara lain demam,

batuk, pilek, sesak napas, dan nyeri tenggorokan. Perbedaan kombinasi gejala tersebut membuat proses identifikasi perlu dilakukan secara cermat oleh tenaga medis.

Organisasi Kesehatan Dunia (WHO, 2014) menempatkan infeksi saluran pernapasan akut sebagai salah satu penyebab utama kesakitan dan kematian akibat penyakit menular di dunia, sehingga deteksi dan penanganan yang tepat waktu menjadi bagian penting dari pelayanan kesehatan. Di Indonesia, ISPA termasuk penyakit yang menjadi perhatian dalam survei kesehatan nasional (Kementerian Kesehatan Republik Indonesia, 2019). Pada fasilitas pelayanan kesehatan tingkat pertama seperti klinik pratama, pasien dengan keluhan pernapasan datang dengan variasi gejala yang luas, sehingga pencatatan gejala yang konsisten dapat menjadi bahan awal untuk analisis data.

Dalam latar belakang penelitian ini, kondisi lingkungan di wilayah Parung Panjang yang memiliki aktivitas kendaraan berat, debu, asap, dan polusi udara disebut sebagai salah satu hal yang perlu diperhatikan karena berhubungan dengan risiko gangguan pernapasan. Di sisi lain, proses identifikasi ISPA di Klinik Pratama Rawat Inap HMS Medika dijelaskan masih mengandalkan pemeriksaan dan analisis tenaga medis. Data pemeriksaan yang tersimpan juga dinilai belum dimanfaatkan secara maksimal untuk menemukan pola yang dapat mendukung prediksi.

Salah satu pendekatan yang dapat digunakan untuk mengelompokkan data berdasarkan pola gejala adalah data mining. Penelitian ini menggunakan algoritma Naive Bayes, yaitu metode klasifikasi probabilistik yang memperkirakan kelas suatu data berdasarkan peluang kemunculan atribut terhadap kelas tersebut. Metode ini dipilih untuk membantu menghasilkan prediksi awal apakah data gejala pasien lebih mengarah pada kelas ISPA atau bukan ISPA.

Tujuan penelitian adalah menerapkan algoritma Naive Bayes dalam prediksi ISPA dan membangun sistem berbasis web yang dapat membantu proses identifikasi awal di Klinik Pratama Rawat Inap HMS Medika. Hasil prediksi tidak dimaksudkan untuk menggantikan pemeriksaan klinis maupun keputusan dokter.

Berdasarkan uraian tersebut, rumusan masalah penelitian ini adalah bagaimana menerapkan algoritma Naive Bayes untuk memprediksi kelas ISPA atau bukan ISPA berdasarkan atribut umur, jenis kelamin, dan gejala pasien, serta bagaimana merancang aplikasi berbasis web yang dapat menampilkan hasil prediksi tersebut secara terstruktur bagi pengguna di klinik. Ruang lingkup penelitian dibatasi pada klasifikasi dua kelas, penggunaan data dummy dan simulasi, serta fungsi prediksi sebagai bagian dari aplikasi pelayanan klinik yang lebih luas.

Kontribusi penelitian ini bersifat praktis dan akademis. Secara praktis, aplikasi yang dirancang memberi gambaran bagaimana data gejala dapat diolah menjadi informasi prediksi awal yang mudah dibaca oleh tenaga medis. Secara akademis, artikel ini mendokumentasikan alur penerapan Naive Bayes pada data gejala ISPA, mulai dari persiapan data, perhitungan peluang, perancangan sistem, hingga rencana pengujian, sekaligus menegaskan batas-batas klaim yang dapat dibuat dari data yang tersedia. Pembahasan disusun dengan memisahkan keberhasilan implementasi fungsi aplikasi dari kinerja klasifikasi yang harus dibuktikan melalui evaluasi terukur.

2. TINJAUAN PUSTAKA

Infeksi saluran pernapasan akut

ISPA adalah kelompok penyakit infeksi yang menyerang bagian saluran pernapasan, baik saluran atas maupun bawah. Manifestasi klinisnya beragam, antara lain demam, batuk, pilek, nyeri tenggorokan, dan sesak napas, dan kombinasi gejala tersebut dapat tumpang tindih dengan penyakit pernapasan lain. Karena itu, penentuan diagnosis pada praktik klinis tetap membutuhkan anamnesis, pemeriksaan fisik, dan penilaian tenaga medis. Sistem prediksi berbasis data hanya dapat berperan sebagai pendukung, yaitu memberi gambaran awal berdasarkan pola gejala yang tercatat.

Data mining dan klasifikasi

Data mining didefinisikan sebagai proses menemukan pola yang bermakna dari sekumpulan data yang besar, dan dipandang sebagai salah satu langkah dalam proses penemuan pengetahuan dari basis data atau knowledge discovery in databases (Fayyad et al., 1996; Han et al., 2012). Rangkaian proses tersebut umumnya mencakup pembersihan data, integrasi, seleksi, transformasi, penambangan pola, evaluasi pola, dan penyajian pengetahuan. Salah satu tugas utama data mining adalah klasifikasi, yaitu membangun model dari data berlabel sehingga model dapat memperkirakan label kelas untuk data baru (Han et al., 2012). Pada penelitian ini, label kelas yang dimaksud adalah ISPA dan bukan ISPA.

Kualitas data sangat memengaruhi hasil klasifikasi. Kotsiantis et al. (2006) membahas bahwa persiapan data pada pembelajaran terbimbing mencakup penanganan nilai yang hilang, pengurangan derau, dan pemilihan atribut yang relevan. Tahap-tahap ini sejalan dengan preprocessing yang dilakukan dalam penelitian ini, yaitu pemeriksaan kelengkapan, penanganan data tidak lengkap, dan pemilihan atribut.

Algoritma Naive Bayes

Naive Bayes adalah pengklasifikasi probabilistik yang berlandaskan Teorema Bayes dan mengasumsikan atribut saling bebas secara kondisional apabila kelas diketahui (Han et al., 2012; Witten et al., 2017). Asumsi tersebut jarang terpenuhi sepenuhnya pada data nyata, tetapi Domingos dan Pazzani (1997) menunjukkan bahwa pengklasifikasi Bayes sederhana dapat tetap menghasilkan klasifikasi yang baik pada banyak kasus meskipun asumsi independensi dilanggar. Rish (2001) melalui studi empiris juga melaporkan bahwa Naive Bayes dapat bekerja baik pada beragam kondisi ketergantungan antaratribut, sedangkan Lewis (1998) mengulas peran asumsi independensi dalam penerapan Naive Bayes pada temu kembali informasi.

Dalam bidang medis, Kononenko (2001) menelaah perkembangan pembelajaran mesin untuk diagnosis medis dan menekankan pentingnya model yang dapat dijelaskan kepada dokter. Naive Bayes relevan dalam konteks ini karena perhitungannya transparan: kontribusi setiap gejala terhadap kelas dapat ditelusuri melalui nilai peluang bersyarat. Selain itu, algoritma ini membutuhkan komputasi yang ringan, sehingga mudah diimplementasikan pada aplikasi web.

Algoritma pembanding

Decision Tree dan Random Forest sering dijadikan pembanding Naive Bayes. Decision Tree membangun model berbentuk pohon keputusan melalui pemisahan data secara bertahap berdasarkan atribut yang paling informatif (Quinlan, 1986), sedangkan Random Forest menggabungkan banyak pohon keputusan yang dilatih pada sampel acak untuk meningkatkan kestabilan prediksi (Breiman, 2001). Kedua algoritma ini disebut dalam saran penelitian lanjutan sebagai pembanding.

Evaluasi kinerja klasifikasi

Kinerja pengklasifikasi biner umumnya dievaluasi dengan confusion matrix yang memuat jumlah true positive (TP), false positive (FP), true negative (TN), dan false negative (FN) (Sokolova & Lapalme, 2009). Dari tabel tersebut dapat dihitung akurasi, presisi, recall, dan F1-score. Akurasi saja dapat menyesatkan ketika proporsi kelas tidak seimbang, sehingga ukuran lain perlu dilaporkan (Hossin & Sulaiman, 2015). Untuk memperkirakan kinerja pada data yang belum dilihat model, validasi silang merupakan metode yang lazim digunakan (Kohavi, 1995), dan analisis ROC dapat melengkapi evaluasi (Fawcett, 2006).

Penelitian terkait

Abdillah dan Dewi (2024) mengembangkan aplikasi berbasis web dengan framework CodeIgniter untuk memprediksi status kegawatan pasien tuberkulosis menggunakan Naive Bayes Classifier di Kota Padang. Penelitian tersebut memiliki kesamaan dengan penelitian ini dalam hal kerangka kerja aplikasi dan algoritma, tetapi berbeda pada penyakit yang diprediksi dan variabel yang dipakai. Penelitian ini memosisikan diri pada kasus ISPA di klinik pratama dengan atribut berupa umur, jenis kelamin, dan lima gejala klinis.

3. METODE

Penelitian berfokus pada penerapan data mining dengan algoritma Naive Bayes serta pembangunan aplikasi prediksi ISPA berbasis web. Lokasi penelitian yang disebutkan dalam skripsi adalah Klinik Pratama Rawat Inap HMS Medika.

Tahapan penelitian

Tahapan penelitian mengikuti alur umum penemuan pengetahuan dari data (Fayyad et al., 1996), yaitu pengumpulan informasi, persiapan data, penerapan algoritma, perancangan dan implementasi sistem, serta pengujian dan evaluasi. Pada tahap awal, wawancara dan observasi dipakai untuk memahami alur pelayanan dan gejala yang dicatat pada pemeriksaan pasien, sedangkan studi literatur dipakai untuk menentukan algoritma dan ukuran evaluasi.

Pengumpulan informasi

Penelitian ini menjelaskan tiga cara pengumpulan informasi, yaitu wawancara dengan tenaga medis atau pihak terkait, observasi terhadap alur pelayanan dan pemeriksaan pasien, serta studi literatur mengenai data mining, Naive Bayes, dan prediksi penyakit ISPA. Batasan penelitian secara eksplisit menyatakan bahwa data yang digunakan berupa data dummy dan simulasi. Dengan demikian, artikel ini tidak menyatakan bahwa seluruh proses pemodelan menggunakan rekam medis pasien nyata.

Data dan atribut

Atribut yang dijelaskan dalam skripsi mencakup umur, jenis kelamin, demam, batuk, pilek, sesak napas, dan nyeri tenggorokan. Kelas keluaran terdiri atas ISPA dan bukan ISPA. Atribut gejala menjadi masukan untuk proses klasifikasi, sedangkan kelas diagnosis digunakan sebagai label dalam data pembelajaran.

Tabel 1 merangkum atribut masukan dan kelas keluaran yang digunakan. Atribut umur bersifat numerik dan dapat ditangani dengan mengelompokkan nilai ke dalam kategori atau memodelkannya dengan distribusi normal (Witten et al., 2017); dokumen sumber belum menjelaskan perlakuan atribut umur secara rinci.

Tabel 1. Atribut masukan dan kelas keluaran

No	Atribut	Peran	Keterangan
1	Umur	Masukan	Karakteristik pasien
2	Jenis kelamin	Masukan	Karakteristik pasien
3	Demam	Masukan	Gejala klinis yang dimasukkan ke sistem
4	Batuk	Masukan	Gejala klinis yang dimasukkan ke sistem
5	Pilek	Masukan	Gejala klinis yang dimasukkan ke sistem
6	Sesak napas	Masukan	Gejala klinis yang dimasukkan ke sistem
7	Nyeri tenggorokan	Masukan	Gejala klinis yang dimasukkan ke sistem
8	Kelas diagnosis	Keluaran (label)	ISPA atau bukan ISPA

Preprocessing data

Sebelum perhitungan dilakukan, data dipersiapkan melalui tahap preprocessing. Tahap ini mencakup pemeriksaan kelengkapan data, penghapusan atau penanganan data yang tidak lengkap, pemilihan atribut yang relevan, serta pengelompokan data berdasarkan kelas. Setelah itu, data disiapkan untuk proses pelatihan dan pengujian. Penelitian ini belum memberikan rincian jumlah baris data, proporsi pembagian data, maupun contoh tabel dataset secara lengkap.

Pemeriksaan kelengkapan bertujuan memastikan setiap catatan memiliki nilai pada seluruh atribut yang dibutuhkan. Catatan yang tidak lengkap dapat dihapus atau dilengkapi sesuai aturan yang ditetapkan, bergantung pada pengaruhnya terhadap proporsi kelas (Kotsiantis et al., 2006). Pemilihan atribut dilakukan untuk mempertahankan atribut yang relevan dengan identifikasi ISPA, dalam hal ini umur, jenis kelamin, dan lima gejala pernapasan. Pengelompokan berdasarkan kelas diperlukan agar peluang prior dan peluang bersyarat dapat dihitung secara terpisah untuk kelas ISPA dan bukan ISPA.

Algoritma Naive Bayes

Naive Bayes merupakan metode klasifikasi yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat saling bebas secara kondisional terhadap kelas. Secara umum, peluang suatu kelas C berdasarkan atribut X dapat dituliskan sebagai:

$$P(C|X) = P(C) \times P(X|C) / P(X)$$

Dalam proses klasifikasi, nilai peluang untuk setiap kelas dihitung. Kelas dengan nilai posterior paling tinggi dipilih sebagai hasil prediksi. Untuk atribut gejala yang bersifat kategorikal, peluang kemunculan setiap nilai atribut pada masing-masing kelas dihitung dari data latih. Nilai prior kelas dihitung dari proporsi jumlah data pada kelas tersebut.

$$P(C|X) \propto P(C) \times \prod P(x_i|C), \quad i = 1, 2, \ldots, n$$

Karena penyebut P(X) bernilai sama untuk semua kelas, perbandingan antarkelas cukup dilakukan pada pembilang, yaitu perkalian peluang prior dengan peluang bersyarat setiap atribut (Han et al., 2012). Pada persamaan tersebut, x_i adalah nilai atribut ke-i dan n adalah jumlah atribut, yaitu tujuh dalam penelitian ini.

Apabila suatu nilai atribut tidak pernah muncul pada kelas tertentu di data latih, peluang bersyaratnya bernilai nol dan menghapus seluruh hasil perkalian. Masalah ini umumnya diatasi dengan pemulusan Laplace (Witten et al., 2017), yaitu menambahkan satu pada setiap hitungan nilai atribut:

$$P(x_i|c) = (n_{ic} + 1) / (n_c + k)$$

dengan n_{ic} adalah jumlah data kelas c yang memiliki nilai x_i, n_c adalah jumlah data pada kelas c, dan k adalah jumlah nilai berbeda yang mungkin dimiliki atribut tersebut. Dokumen sumber tidak menjelaskan secara eksplisit apakah pemulusan Laplace diterapkan pada sistem, sehingga uraian ini disampaikan sebagai landasan teoretis dan bukan sebagai klaim implementasi.

Pembagian data dan evaluasi

Data latih digunakan untuk menghitung peluang prior dan peluang bersyarat, sedangkan data uji digunakan untuk menilai prediksi pada data yang tidak dipakai saat pelatihan. Pembagian data tunggal atau

validasi silang dapat dipilih; Kohavi (1995) merekomendasikan validasi silang berlapis sepuluh lipatan untuk estimasi akurasi. Ukuran evaluasi yang direncanakan adalah confusion matrix dan akurasi, yang dapat dilengkapi presisi, recall, dan F1-score (Sokolova & Lapalme, 2009).

Penelitian ini juga menyebut confusion matrix dan perhitungan akurasi sebagai cara evaluasi hasil klasifikasi. Akan tetapi, angka akurasi dan komponen confusion matrix tidak tersedia dalam bagian hasil yang menjadi dasar penyusunan artikel ini. Nilai tersebut tidak dibuat atau diperkirakan dalam artikel.

Perancangan, implementasi, dan pengujian sistem

Aplikasi dirancang sebagai sistem berbasis web menggunakan PHP, framework CodeIgniter, dan basis data MySQL. Perancangan sistem dalam penelitian ini mencakup pemodelan UML, antara lain Use Case Diagram, Activity Diagram, Sequence Diagram, dan Class Diagram. Pengujian fungsi aplikasi dijelaskan menggunakan Black Box Testing dengan perhatian pada masukan data pasien, masukan gejala, proses perhitungan Naive Bayes, hasil prediksi, dan laporan hasil prediksi.

UML dipakai sebagai bahasa pemodelan untuk menggambarkan kebutuhan fungsional, alur aktivitas, interaksi antarobjek, dan struktur kelas aplikasi (Booch et al., 2005). Black Box Testing dipilih karena menguji perilaku fungsi dari sisi masukan dan keluaran tanpa melihat struktur kode program (Myers et al., 2011; Pressman & Maxim, 2020).

4. HASIL DAN DISKUSI

Analisis sistem dan modul aplikasi

Analisis sistem dalam penelitian ini menggambarkan proses pelayanan klinik yang sebelumnya masih dilakukan secara manual, termasuk pengelolaan pendaftaran, pemeriksaan, pemberian obat, dan pembayaran. Untuk membantu pengelolaan informasi tersebut, aplikasi dirancang memiliki beberapa bagian, seperti login, dashboard, pendaftaran pasien, pemeriksaan, rekam medis, daftar obat, resep, serta pembayaran pasien. Fokus artikel ini tetap pada bagian pengolahan gejala dan prediksi ISPA menggunakan Naive Bayes.

Tabel 2 merangkum modul aplikasi yang disebutkan dalam penelitian beserta fungsi umumnya. Modul prediksi ISPA menjadi fokus artikel ini, sedangkan modul lain mendukung pelayanan klinik secara keseluruhan.

Tabel 2. Modul aplikasi dan fungsi umumnya

No	Modul	Fungsi umum
1	Login	Autentikasi pengguna sebelum mengakses aplikasi
2	Dashboard	Ringkasan informasi dan akses ke modul lain
3	Pendaftaran pasien	Pencatatan data identitas pasien yang berkunjung
4	Pemeriksaan	Pencatatan gejala dan masukan data untuk prediksi ISPA
5	Rekam medis	Penyimpanan riwayat pemeriksaan pasien
6	Daftar obat	Pengelolaan data obat
7	Resep	Pencatatan obat yang diberikan kepada pasien
8	Pembayaran pasien	Pencatatan transaksi pelayanan

Alur prediksi Naive Bayes

Proses prediksi dimulai ketika data gejala pasien dimasukkan ke dalam sistem. Data tersebut kemudian diproses dengan menghitung peluang awal masing-masing kelas, yaitu ISPA dan bukan ISPA. Selanjutnya sistem menghitung peluang setiap atribut gejala terhadap masing-masing kelas. Hasil perhitungan digabungkan untuk memperoleh nilai posterior setiap kelas. Sistem kemudian memilih kelas dengan nilai posterior tertinggi sebagai hasil prediksi.

Alur tersebut memungkinkan data gejala yang dimasukkan diproses secara konsisten berdasarkan pola pada data yang digunakan. Meski demikian, kualitas hasil prediksi sangat bergantung pada kelengkapan, keterwakilan, dan ketepatan data latih. Penggunaan data dummy dan simulasi berarti hasil penerapan perlu diuji kembali menggunakan data yang memadai sebelum digunakan dalam lingkungan pelayanan klinis.

Secara konseptual, alur prediksi dapat diuraikan dalam empat langkah. Pertama, sistem menghitung peluang prior setiap kelas dari proporsi data latih. Kedua, sistem menghitung peluang bersyarat setiap nilai atribut terhadap masing-masing kelas. Ketiga, peluang bersyarat dikalikan dengan peluang prior untuk memperoleh skor posterior setiap kelas. Keempat, kelas dengan skor tertinggi ditampilkan sebagai hasil

prediksi. Karena hanya dua kelas yang digunakan, keputusan akhir pada dasarnya merupakan perbandingan dua skor posterior. Keluaran berupa kelas ISPA atau bukan ISPA ini dapat dibaca langsung oleh pengguna tanpa interpretasi statistik tambahan.

Salah satu keunggulan Naive Bayes dalam konteks klinik adalah keterlacakan hasil. Setiap gejala memberi kontribusi peluang yang dapat diperiksa terhadap kelas yang dipilih, sehingga tenaga medis dapat memahami mengapa suatu data cenderung masuk ke kelas tertentu. Karakteristik ini sejalan dengan kebutuhan model medis yang dapat dijelaskan (Kononenko, 2001). Namun, kemampuan menjelaskan tidak sama dengan ketepatan klasifikasi. Model yang transparan tetap dapat menghasilkan prediksi yang keliru apabila data latihnya tidak representatif.

Implementasi sistem

Sistem dikembangkan dalam bentuk aplikasi web dengan CodeIgniter dan MySQL. Berdasarkan uraian implementasi pada Penelitian ini, aplikasi menyediakan fungsi pengelolaan data pasien dan pelayanan klinik. Pada bagian prediksi, sistem menerima data pasien dan gejala, menjalankan proses perhitungan Naive Bayes, lalu menampilkan hasil klasifikasi. Bentuk aplikasi berbasis web diharapkan memudahkan pengguna dalam memasukkan data dan melihat hasil pengolahan secara terstruktur.

Penggunaan CodeIgniter memisahkan logika aplikasi, tampilan, dan akses data, sehingga modul prediksi dapat dikembangkan terpisah dari modul pendaftaran, rekam medis, obat, resep, dan pembayaran. MySQL digunakan sebagai basis data aplikasi. Kombinasi framework dan algoritma yang sama juga dipakai oleh Abdillah dan Dewi (2024) pada aplikasi prediksi status kegawatan pasien tuberkulosis, tetapi dalam konteks penyakit yang berbeda.

Pengujian sistem

Penelitian ini menjelaskan penggunaan Black Box Testing untuk memeriksa fungsi masukan data pasien, masukan gejala, perhitungan algoritma, keluaran prediksi, dan laporan. Selain itu, evaluasi performa klasifikasi direncanakan menggunakan confusion matrix dan akurasi. Namun, dokumen skripsi yang menjadi sumber artikel ini belum menyajikan tabel hasil pengujian secara rinci, jumlah data uji, nilai true positive, true negative, false positive, false negative, maupun persentase akurasi. Karena itu, belum dapat disampaikan klaim kuantitatif mengenai tingkat keberhasilan model.

Untuk memperjelas ruang lingkup pengujian, Tabel 3 merangkum skenario Black Box yang dirancang berdasarkan fungsi yang disebut dalam penelitian. Tabel ini memuat skenario dan hasil yang diharapkan, bukan hasil pengujian aktual, karena hasil pengujian belum tersedia dalam dokumen sumber.

Tabel 3. Skenario pengujian Black Box yang dirancang

No	Fungsi yang diuji	Skenario	Hasil yang diharapkan
1	Masukan data pasien	Mengisi data pasien dengan data lengkap dan valid	Data tersimpan dan dapat ditampilkan kembali
2	Masukan data pasien	Mengosongkan kolom yang wajib diisi	Sistem menolak masukan dan menampilkan pesan kesalahan
3	Masukan gejala	Memilih kombinasi gejala yang valid	Gejala tersimpan dan siap diproses
4	Proses Naive Bayes	Menjalankan perhitungan pada data gejala yang lengkap	Skor kedua kelas dihitung dan kelas dengan skor tertinggi dipilih
5	Hasil prediksi	Menampilkan keluaran perhitungan	Hasil tampil sebagai ISPA atau bukan ISPA
6	Laporan	Membuka laporan hasil prediksi	Laporan tampil sesuai data yang tersimpan

Teknik Black Box umumnya memakai partisi ekuivalensi dan analisis nilai batas untuk merancang kasus uji (Myers et al., 2011). Pada atribut umur, misalnya, kasus uji perlu mencakup nilai yang sah, nilai kosong, dan nilai di luar rentang yang wajar.

Evaluasi klasifikasi

Tabel 4 menunjukkan struktur confusion matrix dengan ISPA sebagai kelas positif. Tabel ini hanya memperlihatkan struktur dan tidak memuat nilai, karena jumlah data uji dan hasil klasifikasi belum tersedia.

Tabel 4. Struktur confusion matrix

	Prediksi: ISPA	Prediksi: bukan ISPA
Aktual: ISPA	True positive (TP)	False negative (FN)
Aktual: bukan ISPA	False positive (FP)	True negative (TN)

Dari keempat komponen tersebut, ukuran kinerja dihitung dengan rumus berikut (Sokolova & Lapalme, 2009).

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Presisi} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-score} = 2 \times \text{Presisi} \times \text{Recall} / (\text{Presisi} + \text{Recall})$$

Pada skrining penyakit, false negative, yaitu pasien ISPA yang diprediksi bukan ISPA, umumnya lebih berisiko daripada false positive, sehingga recall perlu mendapat perhatian khusus di samping akurasi.

Pada bagian kesimpulan penelitian ini disebutkan bahwa sistem berhasil diterapkan untuk menghasilkan prediksi ISPA atau bukan ISPA berdasarkan atribut pasien. Pernyataan tersebut menggambarkan hasil implementasi sistem, tetapi tidak dapat disamakan dengan bukti bahwa model telah mencapai nilai akurasi tertentu. Perlu dibedakan antara keberhasilan fungsi aplikasi dan kinerja klasifikasi yang harus dibuktikan melalui hasil evaluasi terukur.

Pembahasan

Dibandingkan dengan pengamatan manual, prediksi berbasis data berpotensi menyeragamkan cara gejala diolah. Namun, potensi ini baru dapat dinilai setelah model diuji pada data yang representatif. Penggunaan data dummy dan simulasi membuat pola gejala pada data bersifat buatan, sehingga peluang yang dihitung belum tentu mencerminkan pola pasien di Klinik Pratama Rawat Inap HMS Medika. Karena itu, nilai akurasi yang kelak diperoleh dari data simulasi tidak boleh digeneralisasi ke pasien nyata.

Beberapa faktor perlu diperhatikan pada penelitian lanjutan. Pertama, ketidakseimbangan kelas dapat membuat akurasi tampak tinggi meskipun recall kelas ISPA rendah (Hossin & Sulaiman, 2015). Kedua, gejala seperti demam, batuk, dan pilek juga muncul pada penyakit pernapasan lain, sehingga keterbatasan atribut dapat membuat kelas bukan ISPA sulit dibedakan. Ketiga, asumsi independensi dapat dilanggar karena beberapa gejala cenderung muncul bersamaan, walaupun hal ini tidak otomatis menurunkan kinerja (Domingos & Pazzani, 1997; Rish, 2001).

Perbandingan dengan Decision Tree dan Random Forest, sebagaimana disarankan dalam penelitian, dapat dilakukan pada dataset dan skema pembagian data yang sama agar hasilnya sebanding. Decision Tree menghasilkan aturan yang mudah dibaca (Quinlan, 1986), sedangkan Random Forest cenderung lebih stabil karena menggabungkan banyak pohon (Breiman, 2001). Perbandingan ini memungkinkan pemilihan model dengan keseimbangan antara akurasi dan keterjelasan.

Aspek etika dan hukum juga penting. Data kesehatan termasuk data pribadi yang bersifat spesifik menurut Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi, dan pengelolaan rekam medis diatur dalam Peraturan Menteri Kesehatan Nomor 24 Tahun 2022. Apabila penelitian lanjutan memakai data pasien, penulis perlu memperoleh izin, menjaga kerahasiaan, dan menganonimkan identitas. Hasil prediksi sistem juga harus ditampilkan sebagai informasi pendukung, dengan keputusan klinis tetap berada pada dokter.

Keterbatasan penelitian

Keterbatasan utama yang perlu diperhatikan adalah penggunaan data dummy dan simulasi sebagaimana dinyatakan dalam batasan penelitian. Selain itu, hasil evaluasi klasifikasi belum dilengkapi angka akurasi maupun confusion matrix. Untuk kepentingan publikasi ilmiah, penulis perlu melengkapi jumlah data, sumber dan periode data, proses pembagian data latih dan data uji, hasil pengujian Black Box, serta confusion matrix. Jika penelitian lanjutan menggunakan data pasien, pengelolaan data harus memperhatikan izin, kerahasiaan, dan perlindungan informasi kesehatan.

5. KESIMPULAN

Penelitian ini membahas penerapan algoritma Naive Bayes untuk memprediksi penyakit ISPA berdasarkan data gejala pasien di Klinik Pratama Rawat Inap HMS Medika. Atribut yang digunakan meliputi umur, jenis kelamin, demam, batuk, pilek, sesak napas, dan nyeri tenggorokan, dengan kelas prediksi ISPA atau bukan ISPA. Sistem dirancang berbasis web menggunakan CodeIgniter dan MySQL untuk membantu pengolahan data serta menampilkan hasil prediksi secara terstruktur. Berdasarkan uraian pada penelitian ini,

sistem dapat digunakan untuk menghasilkan prediksi awal dan berpotensi membantu tenaga medis dalam proses identifikasi. Namun, prediksi tersebut hanya berfungsi sebagai alat bantu dan tidak menggantikan diagnosis dokter. Karena data yang digunakan dinyatakan berupa data dummy dan simulasi serta angka evaluasi belum disajikan, tingkat akurasi model belum dapat disimpulkan secara kuantitatif. Pengembangan berikutnya perlu menggunakan dataset yang jelas dan memadai, melaporkan pembagian data latih dan data uji, serta menyertakan confusion matrix dan nilai akurasi. Penelitian juga dapat membandingkan Naive Bayes dengan algoritma lain, seperti Decision Tree dan Random Forest, sebagaimana disarankan dalam skripsi. Selain itu, penelitian selanjutnya disarankan menambahkan atribut klinis lain yang relevan, melibatkan tenaga medis dalam menilai kewajaran hasil prediksi, dan menguji aplikasi pada kondisi penggunaan nyata di klinik setelah data dan izin penggunaannya tersedia.

REFERENCES

1. Indhira, S., Yuhandri, & Hendrik, B. (2024). Penerapan metode Naïve Bayes dalam mengklasifikasi penyakit ISPA di Puskesmas. *Jurnal KomtekInfo*, *11*(4), 197–204. https://doi.org/10.35134/komtekinfo.v11i4.541
2. Booch, G., Rumbaugh, J., & Jacobson, I. (2005). The unified modeling language user guide (2nd ed.). Addison-Wesley.
3. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
4. Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29(2–3), 103–130.
5. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874.
6. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54.
7. Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
8. Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 1–11.
9. Kementerian Kesehatan Republik Indonesia. (2019). Laporan nasional Riskesdas 2018. Badan Penelitian dan Pengembangan Kesehatan.
10. Kementerian Kesehatan Republik Indonesia. (2022). Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis.
11. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), 2, 1137–1143.
12. Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. Artificial Intelligence in Medicine, 23(1), 89–109.
13. Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised learning. International Journal of Computer Science, 1(2), 111–117.
14. Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. Proceedings of the 10th European Conference on Machine Learning (ECML), 4–15.
15. Myers, G. J., Sandler, C., & Badgett, T. (2011). The art of software testing (3rd ed.). John Wiley & Sons.
16. Pressman, R. S., & Maxim, B. R. (2020). Software engineering: A practitioner's approach (9th ed.). McGraw-Hill Education.
17. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.
18. Republik Indonesia. (2022). Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi.
19. Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3(22), 41–46.
20. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437.
21. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.
22. World Health Organization. (2014). Infection prevention and control of epidemic- and pandemic-prone acute respiratory infections in health care. WHO Press.