

BAB. III

METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk membangun model prediksi penyakit malaria pada data pasien Puskesmas Kota Rantauprapat menggunakan algoritma Naïve Bayes. Metodologi yang digunakan disusun dalam beberapa tahapan sebagaimana ditunjukkan dalam gambar alur penelitian.

3.1. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis komputasi. Algoritma Naïve Bayes diterapkan untuk melakukan klasifikasi data pasien guna memprediksi kemungkinan terjangkit malaria berdasarkan atribut gejala dan data pendukung lainnya.

3.2. Alat dan Perangkat Lunak

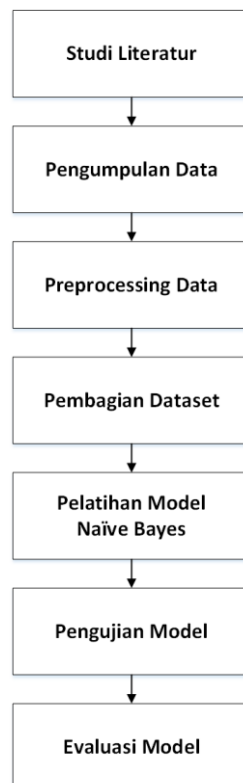
Penelitian ini menggunakan perangkat lunak RapidMiner Studio sebagai alat bantu utama dalam proses klasifikasi dan analisis data. RapidMiner memungkinkan pelaksanaan tahapan machine learning secara visual dan interaktif tanpa memerlukan penulisan kode. Beberapa fitur dan operator utama yang digunakan dalam RapidMiner antara lain:

1. Read Excel: untuk mengimpor data dari file .xlsx atau .csv
2. Replace Missing Values: menangani data kosong atau tidak valid
3. Nominal to Numerical: mengonversi atribut kategorik menjadi numerik
4. Set Role: menetapkan atribut label untuk klasifikasi
5. Split Data: membagi data ke dalam data latih dan data uji

6. Naïve Bayes: untuk membangun model klasifikasi
7. Apply Model: menerapkan model ke data uji
8. Performance (Classification): menghitung metrik evaluasi seperti accuracy, precision, recall, F1-score
9. Confusion Matrix: menampilkan hasil klasifikasi dalam bentuk matriks

3.3. Alur Penelitian

Dengan metodologi yang sistematis ini, diharapkan hasil penelitian dapat memberikan gambaran akurat dan dapat diandalkan dalam memprediksi penyakit malaria. Diagram alur kerja alur penelitian ini diperlihatkan pada Gambar 3.1 di bawah ini.



Gambar 2. 1. Alur Penelitian

Tahapan penelitian ini disusun sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan untuk mengkaji teori-teori yang relevan dengan topik penelitian, termasuk penyakit malaria, algoritma Naïve Bayes, serta penerapan machine learning dalam bidang kesehatan.

2. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Puskesmas Kota Rantauprapat. Data tersebut mencakup atribut-atribut pasien yang berhubungan dengan diagnosis malaria, seperti usia, jenis kelamin, suhu tubuh, gejala, dan hasil pemeriksaan.

3. Preprocessing Data

Tahap ini bertujuan untuk mempersiapkan data agar dapat digunakan dalam pelatihan model. Kegiatan yang dilakukan antara lain:

1. Pembersihan data (menghapus data kosong atau tidak valid)
2. Normalisasi/standarisasi data (jika diperlukan)
3. Transformasi data kategorik menjadi numerik dengan metode label encoding atau one-hot encoding.

4. Pembagian Dataset

Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu:

1. Data training: digunakan untuk melatih model Naïve Bayes.
2. Data testing: digunakan untuk menguji kinerja model.

Rasio pembagian umum yang digunakan adalah 70% data latih dan 30% data uji.

5. Pelatihan Model Naïve Bayes

Model Naïve Bayes dilatih menggunakan data latih. Proses ini melibatkan perhitungan probabilitas masing-masing kelas berdasarkan fitur yang ada dalam dataset.

6. Pengujian Model

Model yang telah dilatih kemudian diuji menggunakan data testing untuk melihat seberapa baik model dapat mengklasifikasikan data baru.

7. Evaluasi Model

Evaluasi dilakukan menggunakan metrik klasifikasi yaitu: Accuracy, Precision, Recall, dan F1-Score. Evaluasi ini bertujuan untuk mengukur seberapa baik model dalam memprediksi penyakit malaria berdasarkan data input yang diberikan.

3.3.1. Studi Literatur

Tahap studi literatur merupakan fondasi utama dalam setiap penelitian ilmiah, termasuk dalam penelitian “Penerapan Algoritma Naive Bayes untuk Prediksi Penyakit Malaria pada Data Pasien Puskesmas Kota Rantauprapat”. Pada tahap ini, peneliti melakukan kajian mendalam terhadap berbagai sumber pustaka yang relevan untuk membangun pemahaman teoritis, metodologis, dan kontekstual yang kuat sebelum melangkah ke tahap berikutnya. Adapun langkah-langkah yang dilakukan pada tahap studi literatur adalah sebagai berikut:

1. Pencarian Sumber Pustaka

Tahap Pencarian Sumber Pustaka merupakan langkah awal yang sangat krusial dalam proses studi literatur. Pada tahap ini, peneliti berupaya

mengumpulkan berbagai referensi ilmiah yang relevan dan kredibel untuk membangun landasan teori serta memahami konteks penelitian secara menyeluruh. Peneliti menggunakan berbagai *platform* dan database ilmiah yang terpercaya, seperti: Google Scholar, PubMed, IEEE Xplore, ScienceDirect, SpringerLink, dan Perpustakaan digital universitas yang menyediakan jurnal, artikel konferensi, buku, dan laporan penelitian yang relevan. Peneliti juga menggunakan fitur filter untuk membatasi hasil pencarian berdasarkan tahun terbit, jenis dokumen, atau relevansi.

2. Seleksi dan Evaluasi Sumber

Tahap Seleksi dan Evaluasi Sumber merupakan langkah krusial setelah proses pencarian sumber pustaka. Pada tahap ini, peneliti melakukan penyaringan dan penilaian terhadap referensi yang telah dikumpulkan untuk memastikan hanya sumber yang relevan, kredibel, dan berkualitas yang digunakan dalam penelitian. Proses ini sangat penting agar landasan teori dan kerangka konseptual yang dibangun benar-benar kuat dan dapat dipertanggungjawabkan.

3. Analisis Isi Literatur

Tahap Analisis Isi Literatur bertujuan untuk menggali, memahami, dan menginterpretasikan isi dari sumber-sumber pustaka yang telah diseleksi dan dievaluasi. Pada tahap ini, peneliti tidak hanya membaca secara sekilas, tetapi melakukan kajian mendalam terhadap isi, konsep, metode, dan hasil penelitian yang relevan untuk membangun landasan teori yang kokoh serta menemukan celah penelitian yang dapat diisi.

4. Sintesis Pengetahuan

Tahap Sintesis Pengetahuan bertujuan untuk mengintegrasikan dan merangkum berbagai informasi, konsep, teori, dan temuan dari sumber-sumber pustaka yang telah dianalisis sebelumnya. Sintesis ini berfungsi sebagai jembatan yang menghubungkan hasil kajian literatur dengan kerangka konseptual dan landasan teori penelitian, sekaligus mengidentifikasi kesenjangan penelitian yang akan diisi oleh studi ini.

3.3.2. Pengumpulan Data

Tahap pengumpulan data merupakan salah satu tahap yang paling penting dalam penelitian ini, karena kualitas dan karakteristik data yang diperoleh akan sangat mempengaruhi akurasi dan keberhasilan model prediksi yang dibangun.

1. Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari Puskesmas Kota Rantauprapat, yang merupakan fasilitas pelayanan kesehatan tingkat pertama di Kabupaten Labuhanbatu, Sumatera Utara. Puskesmas ini aktif menangani berbagai jenis penyakit termasuk malaria, dan memiliki sistem pencatatan dan pelaporan yang relatif terorganisir, baik secara manual maupun digital. Oleh karena itu, Puskesmas Kota Rantauprapat menjadi objek yang strategis untuk dijadikan lokasi studi karena keberadaan kasus malaria yang signifikan serta ketersediaan data riil pasien.

2. Jenis Data

Jenis data yang dikumpulkan adalah data sekunder, yaitu data yang sudah tersedia dan tercatat sebelumnya oleh pihak puskesmas. Data ini berupa:

1. Data riwayat medis pasien yang pernah berkunjung atau dirawat karena dugaan atau diagnosis malaria.
2. Dataset pasien malaria mencakup berbagai atribut yang berkaitan dengan kondisi klinis dan data demografis pasien.
3. Atribut Data

Atribut atau fitur yang dikumpulkan untuk proses klasifikasi antara lain:

Tabel 3. 1. Atribut Data

Atribut	Keterangan
Usia	Umur pasien dalam tahun
Jenis Kelamin	Laki-laki/Perempuan
Suhu Tubuh	Dicatat saat pasien datang
Demam	Ada/Tidak
Menggigil	Ya/Tidak
Mual/Muntah	Ya/Tidak
Sakit Kepala	Ya/Tidak
Hasil Diagnosa Malaria	Positif/Negatif (Label Klasifikasi)

Atribut-atribut tersebut dipilih berdasarkan literatur dan data klinis yang umum digunakan dalam diagnosis malaria.

4. Prosedur Pengumpulan Data

Langkah-langkah pengumpulan data dilakukan sebagai berikut:

a. Permohonan dan Perizinan

Peneliti mengajukan izin penelitian kepada pihak Puskesmas dan Dinas Kesehatan Kabupaten Labuhanbatu. Data yang diperoleh tidak memuat identitas pribadi pasien (anonim) untuk menjaga privasi dan etika penelitian.

b. Seleksi dan Penyaringan Data

Data yang dikumpulkan diseleksi untuk memastikan hanya mencakup kasus yang relevan dengan studi malaria. Data yang tidak lengkap atau tidak sesuai dengan variabel penelitian dieliminasi atau dikategorikan ulang.

c. Dokumentasi dan Konversi Data

Jika data masih berbentuk fisik/manual, dilakukan digitalisasi menggunakan Microsoft Excel atau Google Spreadsheet. Selanjutnya, data dikonversi ke dalam format '.csv' atau '.xlsx' untuk dianalisis menggunakan RapidMiner.

5. Konteks Lokal dan Kebutuhan Data

Puskesmas Kota Rantauprapat memiliki peran penting dalam pengendalian penyakit menular seperti malaria, terutama karena tingginya potensi lingkungan yang mendukung penyebaran nyamuk Anopheles. Penggunaan data lokal memperkuat validitas model yang dikembangkan.

6. Tantangan Pengumpulan Data

Beberapa tantangan yang mungkin dihadapi selama proses pengumpulan data antara lain:

1. Kelengkapan data yang bervariasi
2. Format data yang tidak seragam
3. Verifikasi keabsahan data yang membutuhkan waktu

7. Etika dan Kerahasiaan Data

Penelitian ini menjunjung tinggi prinsip etika, dengan memastikan bahwa data pasien telah dianonimkan dan tidak memuat informasi pribadi.

Penggunaan data dilakukan sesuai standar etik dan ketentuan yang berlaku..

3.3.3. Preprocessing Data

Tahap *preprocessing* data merupakan proses penting dalam data mining dan machine learning. Preprocessing bertujuan untuk membersihkan dan mempersiapkan data agar layak digunakan dalam pelatihan model. Pada penelitian ini, preprocessing dilakukan terhadap data pasien malaria dari Puskesmas Kota Rantauprapat yang telah dikumpulkan pada tahap sebelumnya.

1. Pembersihan Data

Langkah awal dalam preprocessing adalah membersihkan data dari kesalahan atau ketidaksesuaian. Beberapa data memiliki nilai kosong (missing values), kesalahan pengetikan, atau data duplikat. Penanganan dilakukan dengan cara:

1. Menghapus baris data yang kosong seluruhnya atau sebagian jika tidak dapat diperbaiki.
2. Mengisi nilai kosong dengan metode statistik (seperti mean, modus, atau interpolasi) jika relevan.
3. Menghapus data duplikat berdasarkan ID pasien atau kombinasi atribut identik.

2. Transformasi Data

Data yang digunakan dalam penelitian ini terdiri dari atribut numerik dan kategorik. Untuk dapat digunakan dalam model Naïve Bayes, atribut kategorik perlu dikonversi menjadi bentuk numerik. Beberapa langkah transformasi yang dilakukan:

1. Label encoding: Mengonversi atribut seperti jenis kelamin (Laki-laki = 0, Perempuan = 1).
2. Binary encoding: Gejala seperti demam, mual, atau menggigil dikonversi ke 1 (Ya) dan 0 (Tidak).
3. Normalisasi suhu tubuh ke dalam rentang tertentu jika diperlukan.

3. Seleksi Fitur

Seleksi fitur adalah proses memilih atribut yang paling relevan untuk digunakan dalam pelatihan model. Atribut yang tidak berpengaruh signifikan terhadap hasil klasifikasi dapat dieliminasi. Seleksi dilakukan berdasarkan:

1. Relevansi atribut terhadap label (hasil diagnosa)
2. Korelasi antar atribut
3. Pengalaman klinis dan referensi penelitian terdahulu

4. Penyeimbangan Kelas (*Balancing Data*)

Jika dalam dataset jumlah pasien yang positif malaria jauh lebih sedikit atau lebih banyak dari yang negatif, maka dilakukan penyeimbangan kelas. Ketimpangan ini dapat menyebabkan bias dalam model. Teknik yang digunakan antara lain:

1. Oversampling: menambahkan salinan data dari kelas minoritas
 2. Undersampling: mengurangi data dari kelas mayoritas
 3. SMOTE: metode sintetik untuk menyeimbangkan jumlah kelas secara algoritmik
5. Hasil Akhir Preprocessing

Setelah melalui seluruh tahapan preprocessing, data menjadi lebih bersih, seragam, dan siap untuk dibagi menjadi data latih dan data uji. Dataset telah dikonversi ke dalam format yang dapat diproses oleh algoritma Naïve Bayes menggunakan RapidMiner.

3.3.4. Pembagian Dataset

Setelah data selesai melalui tahap preprocessing, langkah selanjutnya adalah melakukan pembagian dataset. Pembagian dataset bertujuan untuk memisahkan data yang akan digunakan untuk melatih model (training) dan data yang akan digunakan untuk menguji kinerja model (testing). Tujuan utama dari pembagian dataset adalah untuk menghindari overfitting, yaitu kondisi ketika model terlalu baik dalam mengenali data latih namun gagal menggeneralisasi terhadap data baru. Dengan memisahkan data latih dan data uji, evaluasi terhadap performa model dapat dilakukan secara objektif.

1. Rasio Pembagian

Dalam penelitian ini, data dibagi menjadi dua bagian dengan rasio sebagai berikut: 70% data digunakan untuk proses pelatihan model (training set). 30% data digunakan untuk proses pengujian model (*testing set*).

Pembagian ini dilakukan secara acak (random sampling) untuk memastikan bahwa data dalam kedua subset terdistribusi secara representatif.

2. Validasi Hasil Pembagian

Setelah pembagian, dilakukan pengecekan terhadap distribusi label (positif dan negatif malaria) pada kedua subset untuk memastikan bahwa tidak terjadi ketidakseimbangan kelas yang signifikan. Hal ini penting agar model tidak bias hanya terhadap salah satu kelas.

3. Hasil Pembagian Dataset

Hasil dari proses ini adalah dua set data: Data latih yang digunakan untuk melatih model Naïve Bayes. Data uji yang digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data baru yang belum dikenali.

3.3.5. Pelatihan Model Naïve Bayes

Tahap pelatihan model merupakan inti dari proses machine learning, di mana algoritma digunakan untuk mempelajari pola dari data latih yang telah disiapkan sebelumnya. Pada penelitian ini, algoritma yang digunakan adalah Naïve Bayes, yang termasuk dalam kelompok supervised learning dan berbasis probabilistik. Naïve Bayes adalah algoritma klasifikasi yang menggunakan prinsip Teorema Bayes dengan asumsi independensi antar fitur. Model ini memprediksi probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan frekuensi nilai atribut pada data latih. Naïve Bayes cocok digunakan dalam prediksi medis karena mampu menangani dataset kecil dengan efisien dan cepat.

1. Persiapan Data Latih

Sebelum model dilatih, data latih dipastikan telah:

1. Bersih dari data yang hilang atau tidak valid
2. Telah melalui proses encoding fitur kategorik menjadi numerik
3. Memiliki distribusi kelas yang seimbang antara positif dan negatif malaria

2. Output Pelatihan

Setelah proses pelatihan, model menghasilkan nilai-nilai probabilitas untuk setiap kelas pada setiap data input. Model ini kemudian siap digunakan untuk memprediksi kelas dari data uji pada tahap selanjutnya. Output model meliputi: Probabilitas masing-masing kelas, Model klasifikasi siap pakai, dan Parameter statistik seperti mean dan standar deviasi untuk tiap fitur dan kelas.

3.3.6. Pengujian Model

Setelah proses pelatihan selesai dilakukan, langkah selanjutnya adalah melakukan pengujian terhadap model yang telah dibangun. Tujuan dari tahap ini adalah untuk mengukur seberapa baik model Naïve Bayes mampu mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Pengujian dilakukan untuk mengevaluasi performa model dalam memprediksi label kelas berdasarkan data uji. Pengujian ini bertujuan untuk memastikan bahwa model tidak hanya bekerja baik pada data latih, tetapi juga pada data uji yang merepresentasikan kasus nyata yang lebih umum.

Model yang telah dilatih kemudian digunakan untuk memprediksi label dari data uji. Hasil pengujian akan digunakan untuk menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Selain itu, akan dibuat confusion matrix untuk mengetahui distribusi prediksi terhadap kelas aktual secara rinci. Validasi hasil pengujian dilakukan dengan membandingkan prediksi model terhadap nilai aktual. Pengujian ini memberikan gambaran tentang seberapa baik model dapat menggeneralisasi terhadap data baru. Jika hasil pengujian menunjukkan performa yang tinggi, maka model dianggap layak untuk digunakan sebagai alat bantu prediksi penyakit malaria.

3.3.7. Evaluasi Model

Setelah dilakukan pengujian model terhadap data uji, langkah terakhir dalam proses pembangunan model adalah melakukan evaluasi terhadap kinerja model. Evaluasi bertujuan untuk mengetahui seberapa baik model dalam melakukan klasifikasi terhadap data baru, khususnya dalam memprediksi penyakit malaria.

1. Metode Evaluasi

Evaluasi model dilakukan menggunakan metrik yang umum digunakan dalam klasifikasi biner, yaitu:

1. Accuracy: Mengukur proporsi prediksi yang benar dari seluruh data uji.
2. Precision: Mengukur seberapa tepat model dalam memprediksi kelas positif.
3. Recall (Sensitivity): Mengukur kemampuan model dalam menemukan seluruh kasus positif.

4. F1-Score: Rata-rata harmonis antara precision dan recall, berguna ketika data tidak seimbang.

2. Confusion Matrix

Confusion Matrix adalah matriks 2x2 yang menunjukkan jumlah prediksi benar dan salah yang dilakukan oleh model. Elemen dari confusion matrix adalah:

1. True Positive (TP): Data positif yang diprediksi positif.
2. True Negative (TN): Data negatif yang diprediksi negatif.
3. False Positive (FP): Data negatif yang diprediksi positif.
4. False Negative (FN): Data positif yang diprediksi negatif.

3. Perhitungan Metrik

Berikut adalah rumus perhitungan masing-masing metrik:

1. $\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$
2. $\text{Precision} = TP / (TP + FP)$
3. $\text{Recall} = TP / (TP + FN)$
4. $\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

4. Interpretasi Hasil

Nilai-nilai dari metrik evaluasi digunakan untuk menilai seberapa baik model Naïve Bayes bekerja dalam klasifikasi penyakit malaria. Jika nilai accuracy dan F1-Score tinggi, maka dapat disimpulkan bahwa model memiliki performa yang baik dan dapat digunakan sebagai alat bantu diagnosis awal di Puskesmas Kota Rantauprapat