



BAB III

METODE PENELITIAN

3.1 Pendahuluan

Bab ini menjelaskan langkah-langkah penelitian yang dilakukan untuk mengklasifikasi penyakit pasien rawat jalan di Puskesmas menggunakan metode *K-Nearest Neighbors* (KNN). Metodologi penelitian mencakup alur penelitian, jenis dan sumber data, teknik pengumpulan data, proses pengolahan data, metode analisis, serta alat yang digunakan dalam penelitian. Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining. Teknik yang digunakan adalah klasifikasi, dengan algoritma *K-Nearest Neighbors* (KNN) untuk memodelkan pola kunjungan pasien berdasarkan atribut-atribut tertentu sehingga dapat memprediksi jenis penyakit rawat jalan.

3.2 kerangka penelitian

Kerangka penelitian dirancang memberikan panduan yang sistematis dan terstruktur dalam pelaksanaan penelitian, Berikut adalah kerangka dalam penelitian ini :



Gambar 1 Kerangka Penelitian

3.3 Data Pasien Rawat Jalan

Pengumpulan data dilakukan untuk memperoleh informasi mendalam terkait kebutuhan jenis data yang diperlukan yaitu Mengambil data historis kunjungan pasien rawat jalan dari Puskesmas, termasuk atribut seperti umur, jenis kelamin, waktu kunjungan, poli tujuan, dan diagnosa.

3.4 Preprocessing Data

Preprocessing data merupakan tahap awal dalam penelitian yang bertujuan untuk menyiapkan data pasien rawat jalan agar siap digunakan dalam proses klasifikasi menggunakan metode *K-Nearest Neighbors* (KNN). Tahap ini dilakukan untuk meningkatkan kualitas data sehingga hasil klasifikasi menjadi lebih akurat[11]



B. Proses preprocessing meliputi data cleaning, yaitu pembersihan data dari nilai yang hilang, data duplikat, dan kesalahan penulisan; data transformation, yaitu mengubah data kategorikal menjadi bentuk numerik; serta data normalization, yaitu menyamakan skala data agar perhitungan jarak pada metode KNN dapat dilakukan secara optimal.

Tahap data cleaning dilakukan dengan membersihkan data dari nilai yang hilang (*missing value*), data yang tidak konsisten, serta data yang duplikat[12]. Data yang tidak lengkap dapat dihapus atau diperbaiki agar tidak mempengaruhi hasil analisis. Selain itu, kesalahan penulisan pada data pasien juga diperbaiki untuk menjaga keakuratan data.

Tahap data transformation, yaitu mengubah data ke dalam bentuk yang sesuai untuk proses klasifikasi. Misalnya, data kategorikal seperti jenis kelamin atau gejala diubah menjadi bentuk numerik agar dapat diproses oleh metode KNN. Contohnya, “Laki-laki” diubah menjadi 1 dan “Perempuan” menjadi 0, atau gejala “Ya” menjadi 1 dan “Tidak” menjadi 0.

Tahap data normalization, yaitu proses penyetaraan skala data numerik agar berada dalam rentang yang sama. Hal ini penting karena metode KNN menggunakan perhitungan jarak, sehingga perbedaan skala antar atribut dapat mempengaruhi hasil klasifikasi[13]. Normalisasi dapat dilakukan menggunakan metode seperti Min-Max Normalization.

3.5 Normalisasi Data

Normalisasi data merupakan proses penyesuaian skala nilai pada atribut numerik agar berada dalam rentang yang sama. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, normalisasi sangat penting karena metode KNN menggunakan perhitungan jarak antar data, sehingga perbedaan skala antar atribut dapat mempengaruhi hasil klasifikasi[14]. Atribut seperti usia, suhu tubuh, dan tekanan darah memiliki rentang nilai yang berbeda, sehingga perlu dinormalisasi agar tidak mendominasi perhitungan jarak. Metode normalisasi yang digunakan dalam penelitian ini adalah Min-Max Normalization, dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

di mana:

XXX = nilai data asli

XminX_{min}Xmin = nilai minimum

XmaxX_{max}Xmax = nilai maksimum

X'X'X' = nilai hasil normalisasi

Dengan dilakukan normalisasi, setiap atribut memiliki kontribusi yang seimbang dalam perhitungan jarak, sehingga dapat meningkatkan akurasi hasil klasifikasi menggunakan metode KNN.



3.6 Pembagian Data

Pembagian data merupakan tahap dalam penelitian yang bertujuan untuk memisahkan dataset menjadi data latih (training data) dan data uji (testing data)[15]. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, pembagian data dilakukan untuk menguji kemampuan metode KNN dalam mengklasifikasikan data pasien secara akurat.

Data dibagi secara stratified menjadi tiga bagian agar proporsi tiap penyakit terjaga: 70% train, 15% test. Pembagian ini penting agar evaluasi bersifat objektif data test sama sekali tidak dilibatkan pada tahap pemilihan parameter. Peran masing-masing subset dirangkum pada tabel 3.1

Tabel 3.1 Pembagian data dan fungsinya

Subset	Jumlah data	Fungsi
Train	700	Melatih model KNN
Validation	150	Memilih nilai k terbaik
Test	150	Evaluasi akhir yang objektif

3.7 Proses KNN

Proses K-Nearest Neighbors (KNN) merupakan tahap utama dalam penelitian ini, yaitu menghitung jarak antara data uji dengan data latih untuk menentukan kedekatan antar data. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan



Menggunakan Metode K-Nearest Neighbors (KNN)”, perhitungan jarak digunakan untuk mengetahui tingkat kemiripan kondisi pasien berdasarkan atribut yang dimiliki.

Metode yang digunakan untuk menghitung jarak adalah Euclidean Distance, yaitu metode yang mengukur jarak antara dua titik dalam ruang multidimensi. Rumus Euclidean Distance adalah sebagai berikut:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

di mana:

d = jarak antara data uji dan data latih

x_{ix} = nilai atribut ke- i pada data uji

y_{iy} = nilai atribut ke- i pada data latih

n = jumlah atribut

Setelah seluruh jarak dihitung, langkah selanjutnya adalah mengurutkan jarak dari yang terkecil hingga terbesar untuk menentukan K tetangga terdekat. Data dengan jarak paling kecil dianggap memiliki kemiripan tertinggi dan akan digunakan dalam proses penentuan kelas penyakit pasien.

Dengan menggunakan perhitungan jarak ini, metode KNN dapat mengklasifikasikan data pasien baru berdasarkan kemiripan dengan data pasien sebelumnya.



3.8 Penentuan K Tetangga Terdekat

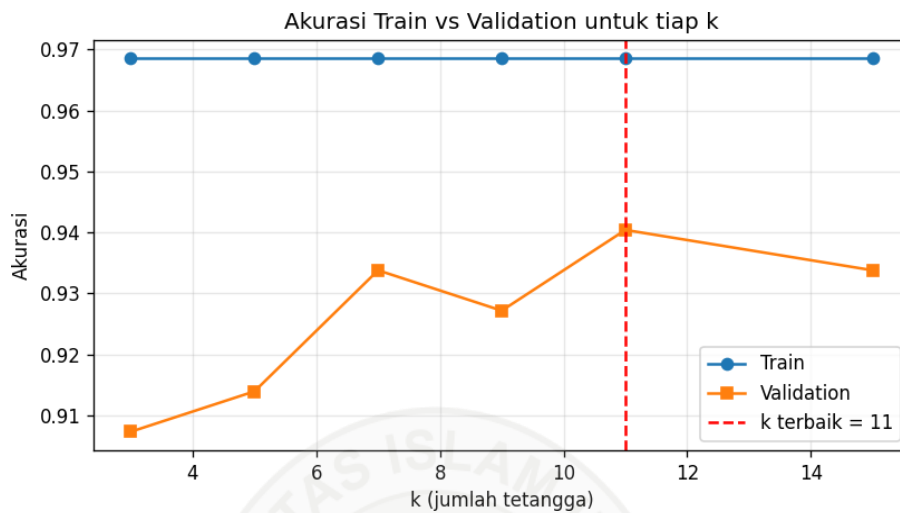
Penentuan K tetangga terdekat merupakan tahap lanjutan dalam metode K-Nearest Neighbors (KNN) setelah proses perhitungan jarak dilakukan. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, nilai K menunjukkan jumlah data tetangga terdekat yang digunakan sebagai acuan dalam menentukan kelas suatu data uji.

Nilai K terbaik ditentukan berdasarkan akurasi tertinggi pada data validation, dengan kandidat K=3, 5, 9, 11, dan 15. Akurasi pada data train data juga dipantau untuk mendeteksi overviting. hasilnya dapat dilihat pada tabel 3.2 dan gambar 3.1

Tabel 3.2 Akurasi train dan validation untuk tiap nilai k

Nilai k	Akurasi Train	Akurasi Validation	Keterangan
3	0,969	0,907	
5	0,969	0,914	
7	0,969	0,934	
9	0,969	0,927	
11	0,969	0,940	Terbaik
15	0,969	0,934	





Gambar 3 Kurva akurasi train vs validation terhadap nilai k

Berdasarkan Tabel 3.2, nilai $K=11$ memberikan akurasi validation tertinggi (0,940) dan dipilih sebagai parameter model akhir. Setelah K terpilih, model final dilatih ulang menggunakan gabungan data train+validation, lalu diuji pada data test

3.9 Evaluasi Model

Evaluasi model merupakan tahap untuk mengukur kinerja metode K-Nearest Neighbors (KNN) dalam mengklasifikasikan penyakit pasien rawat jalan. Pada penelitian dengan judul “Klasifikasi Penyakit Pasien Rawat Jalan Menggunakan Metode K-Nearest Neighbors (KNN)”, evaluasi dilakukan menggunakan akurasi dan confusion matrix.

Akurasi digunakan untuk mengetahui tingkat ketepatan model dalam mengklasifikasikan data. Nilai akurasi diperoleh dari perbandingan antara jumlah data yang diklasifikasikan dengan benar terhadap seluruh data yang diuji. Rumus akurasi adalah sebagai berikut:





$$Akurasi = \frac{Jumlah\ data\ benar}{Total\ data\ uji} \times 100\%$$

Semakin tinggi nilai akurasi, maka semakin baik kinerja model dalam melakukan klasifikasi.

Selain itu, digunakan juga confusion matrix, yaitu tabel yang digunakan untuk mengevaluasi hasil klasifikasi dengan membandingkan antara nilai aktual dan hasil prediksi. Confusion matrix terdiri dari empat komponen, yaitu:

True Positive (TP): data yang diprediksi benar sesuai kelas positif

True Negative (TN): data yang diprediksi benar sesuai kelas negatif

False Positive (FP): data yang diprediksi positif tetapi sebenarnya negatif

False Negative (FN): data yang diprediksi negatif tetapi sebenarnya positif

Melalui confusion matrix, dapat diketahui performa model secara lebih rinci, tidak hanya dari sisi akurasi tetapi juga kesalahan klasifikasi yang terjadi.

Dengan menggunakan akurasi dan confusion matrix, diharapkan dapat diperoleh gambaran yang jelas mengenai tingkat keberhasilan metode KNN dalam klasifikasi penyakit pasien rawat jalan.

Hasil & kesimpulan

Pada tahap hasil, penelitian ini menyajikan output dari penerapan metode K-Nearest Neighbors (KNN) dalam mengklasifikasikan penyakit pasien rawat jalan. Hasil yang diperoleh berupa kelas penyakit dari data uji, yang ditentukan berdasarkan



kemiripan dengan data latih menggunakan perhitungan jarak. Selain itu, ditampilkan juga hasil evaluasi model menggunakan akurasi dan confusion matrix untuk mengetahui tingkat kinerja metode KNN. Dari hasil pengujian, dapat diketahui seberapa baik metode KNN dalam mengklasifikasikan penyakit berdasarkan data pasien yang digunakan.

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Nearest Neighbors (KNN) mampu digunakan untuk mengklasifikasikan penyakit pasien rawat jalan dengan tingkat akurasi tertentu, tergantung pada nilai K yang digunakan dan kualitas data. Metode ini bekerja dengan baik dalam mengidentifikasi kemiripan antar data pasien, sehingga dapat membantu dalam proses klasifikasi penyakit. Dengan demikian, metode KNN dapat dijadikan sebagai sistem pendukung keputusan dalam bidang kesehatan, khususnya dalam membantu tenaga medis dalam mengelompokkan penyakit pasien secara lebih cepat dan sistematis.

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.